JMIR Human Factors

Impact Factor (2023): 2.6
Volume 6 (2019), Issue 2 ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

Contents

Original Papers

Original Paper

A Mobile Phone App for the Self-Management of Pediatric Concussion: Development and Usability Testing

Harminder Sandhu[1], BEng, MHSc; Katherine Wilson[2], MSc; Nick Reed[2,3,4], PhD; Alex Mihailidis[1,4,5], PhD, PEng

[1]Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[2]Concussion Centre, Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada

[3]Rehabilitation Sciences Institute, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[4]Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[5]Toronto Rehabilitation Institute, University Health Network, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Nick Reed, PhD
Concussion Centre
Bloorview Research Institute
Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Road
Toronto, ON, M4G 1R8
Canada
Phone: 1 416 425 6220 ext 3861
Fax: 1 416 425 1634
Email: nreed@hollandbloorview.ca

Abstract

Background: Concussion is a common injury among Canadian children and adolescents that leads to a range of neurobehavioral deficits. However, noticeable gaps continue to exist in the management of pediatric concussion, with poor health outcomes associated with the inadequate application of best practice guidelines.

Objective: The aim of this study was to describe the development and assess the usability of a mobile phone app to aid youth in the self-management of concussion. A secondary objective was to assess the usefulness of the app.

Methods: An agile user-centered design approach was used to develop the technology, followed by a formative lab-based usability study for assessment and improvement proposals. Youths aged 10 to 18 years with a history of concussion and health care professionals involved in concussion management were recruited. This study included participants performing 12 tasks with the mobile phone app while using the *think aloud* protocol and the administration of the System Usability Scale (SUS), posttest questionnaire, and a semistructured interview.

Results: A mobile phone app prototype called *NeuroCare*, an easily accessible pediatric concussion management intervention that provides easy access to expert-informed concussion management strategies and helps guide youth in self-managing and tracking their concussion recovery, was developed. A total of 7 youths aged between 10 and 18 years with a history of concussion and 7 health care professionals were recruited. The mean SUS score was 81.9, mean task success rates were greater than 90% for 92% (11/12) of the tasks, 92% (11/12) of tasks had a total error frequency of less than 11 errors, and mean task completion times were less than 2 min for 100% of the tasks.

Conclusions: Results suggest that participants rated this app as highly usable, acceptable to users, and that it may be useful in helping youth self-manage concussion.

(JMIR Hum Factors 2019;6(2):e12135) doi:10.2196/12135

KEYWORDS

brain concussion; safety; pediatrics; youth; children; self-management; mild traumatic brain injury; mobile apps; mobile health

Introduction

Background

Concussion is a common injury [1,2] (200 per 100,000 [3]) among Canadian children and adolescents that leads to a range of neurobehavioral deficits including combinations of somatic, physical, cognitive, and emotional and behavioral symptoms [4]. These postconcussion symptoms can have a significant impact on the functional participation of youth in daily activities, such as sports, school, as well as family and social activities [3,5,6]. There is a lack of evidence-based interventions for the management of pediatric concussion [6,7], but consistent application of best practice guidelines may help reduce the impact of concussion and persistent postconcussion symptoms [8]. However, noticeable gaps continue to exist in concussion management with inadequate application of best practice guidelines, and there is growing evidence demonstrating both knowledge and practice gaps in concussion management [8-11]. Consequently, individuals may receive inconsistent and incomplete messages regarding the best strategies to manage concussion, which could lead to poor health outcomes.

Concussion & You

Concussion & You is an evidence-informed self-management education program for concussed youth and their families [12]. It features a concussion curriculum based on best evidence and expert opinion and is integrated within a self-management framework. *Concussion & You* aims to provide evidence-informed best practice guidance regarding concussion recovery throughout the entire recovery process and enable participants to build an idiosyncratic concussion recovery toolkit using the practical concussion management strategies provided by the program for the management of return to school and play, sleep, nutrition, relaxation, and energy conservation that the youth can access throughout their recovery [12,13]. The feasibility of this program was validated in a pilot study that led to an increase in patients' knowledge regarding concussion and concussion management strategies after intervention [12]. Youths and their families are able to implement strategies into their daily routines with the use of supplied daily planners, activity logs, and postconcussion symptom scales; results from the postsession survey indicated that these tangible tools positively affect participants' recovery [12]. This program addresses many gaps in concussion management, which can significantly improve the quality of life and outcomes of concussed youth, but it currently relies on in-person interaction and only at 1 time point with no additional follow-up or support.

Mobile Health

Many clinical researchers have begun harnessing technology to develop innovative approaches that hold great promise for enhancing the accessibility and quality of care [14,15]. Mobile health (mHealth) technologies, such as mobile phones, are well suited to serve as platforms for the self-management of health conditions as they are ubiquitous, have great computational capabilities, and are commonly carried on the person [13,16]. In addition, mHealth technologies can facilitate access to self-monitoring resources, time-sensitive health information, prompts, reminders, and personalized self-management tools in real time [13,16]. Mobile phones are ubiquitous in the lives of youth [17,18], so interventions using mobile technology may provide important and innovative opportunities for engaging youth in and improving health-related self-management skills and behaviors [17,19]. We scanned app stores for concussion-related apps and found that concussion-related apps exist, but they are primarily focused on concussion identification, diagnosis, and general information about concussion symptoms and recovery recommendations. These apps are not focused on concussion self-management, are not specific to pediatric concussion, do not emphasize strategies or the provision of tools to promote recovery, and do not allow tracking of personal information or recovery progress. Furthermore, most apps have not been validated in the peer-reviewed literature, and thus, their efficacy in helping youth manage concussion is unclear. Some efforts have been made, as indicated in the literature, to develop and evaluate apps for pediatric concussion management. For example, a study [20] evaluated the effects of a gamified mobile phone app in promoting health management in teenagers with persistent postconcussion symptoms, and it showed promising initial results for the use of mobile phone apps for the management of postconcussion syndrome [20]. However, this app focuses on improving the management of postconcussion syndrome that is experienced by only a subset of all concussed youth [21], and the app does not provide guidance throughout the entire recovery process, thus missing the opportunity to be preventative and guide youth from the onset of injury. SMART, another app developed for pediatric concussion management, is a Web-based educational and self-management program; its initial results show promise for the use of apps for pediatric concussion management [22-24]. However, a usability study of this app identified that some users felt the time and reading required to complete the program would be too difficult for children to comprehend and complete [22-24]. This may suggest the program requires a considerable amount of physical and cognitive effort to use, and the safety with which this app can be used by concussed youth is unclear and should be evaluated in addition to evaluating its efficacy. In addition, the app focuses on managing and tracking symptoms, instead of empowering or enabling the user to implement specific concussion management strategies; Zasler et al discussed that if symptoms persist, then focusing on symptoms might be counterproductive [25]. A technical limitation of Web-based apps is that they require an internet connection, which may not always be available or reliable, limiting access to individuals who have reliable access to the internet. In contrast to Web-based or HTML apps, native apps offer robust offline functionality, which is preferred for mHealth tools targeting individuals who may live in rural areas with poor internet connectivity or who do not have access to the internet [26,27]. Native apps also provide a richer user experience and better and more innovative capabilities than Web-based apps [26,27]. There is a need to develop and evaluate tools that are easily accessible to youth, guide them throughout their concussion recovery, educate them on and assist them in implementing best practice concussion management strategies, and that require minimal engagement by the user to ensure safety.

Usability

Technologies with inadequate consideration of the needs of the intended users are difficult to learn, and these will be misused or underutilized and will ultimately fail to accomplish objectives originally set out [28]. Usability studies are commonly used to evaluate mHealth technologies [29-31], and they focus on measurable user performance and preference metrics. Usability is defined 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 [32,33]. Having a user perform a set of tasks that relate to product features and are representative of the tasks that the user may use the technology for is an excellent way to determine the usability of a feature or feature workflow [34,35]. It is important to perform a usability study for an mHealth technology with prospective end users to effectively determine how well the target audience interacts and relates to a technology.

The objective of this research was to develop and evaluate the usability of a mobile phone app that will help enable youth to better self-manage concussion by providing easy access to expert-informed concussion management information and strategies and a tool that will guide youth in self-managing and tracking concussion recovery. Overall, this app is expected to improve the quality of life of youth who have experienced a concussion by providing recovery support to enable safe return to daily activities of meaning and importance (eg, school, family, social activities, and sport activities).

Methods

Overview of Application, Prototype Design, and Development

A user-centered design approach and Agile development methods were used to design and develop the *NeuroCare* mobile phone app to ensure that it was useful and usable for the end-user population. The design, development, and improvements to the prototype were carried out using an iterative and incremental development (IID) approach [36] with the support of the design team. The design team consisted of the key stakeholders and user proxies that included health care professionals, concussion experts, business personnel, and brain injury researchers who were iteratively involved in the design of this technology, that is, assisted in identifying the end users, creating a target user group profile, creating a persona, identifying design requirements and design principles, identifying mobile app's features and functions, and identifying app content and design.

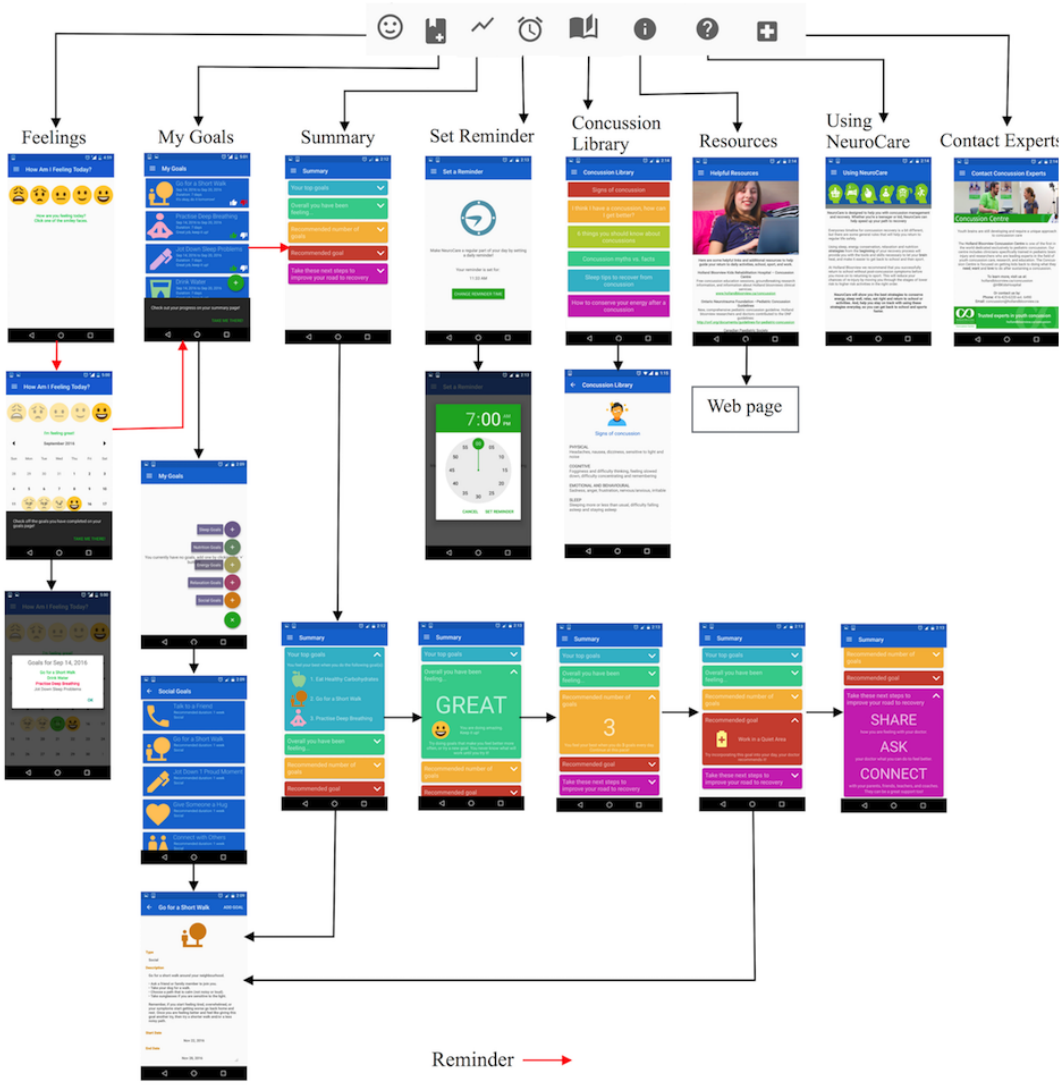
The design team identified that the primary end users of this technology are concussed youth aged between 10 and 18 years, and the secondary end users are the health care professionals who are involved in concussion assessment and management. Concussed youth can use this app to better self-manage their concussion through the implementation of evidence-informed concussion management strategies and progress tracking. Health care professionals can direct their youth clients/patients to the app and work with the youth to review concussion recovery progress and provide better direction and support. Both end users will use this app to improve communication between each other, which has shown to improve patient health outcomes, specifically emotional health and symptom resolution [37]. Youth may gain access to the app through Web-based app stores. The app can be used to support concussion recovery from the time of injury through to recovery and may be the most beneficial once the user begins to reintroduce daily activities at a gradual pace so that these do not provoke symptoms. Youth are instructed to use the app daily throughout their concussion recovery and share their progress (eg, how they have been feeling and the strategies implemented) with their health care professional to receive further support and feedback in managing their concussion (eg, advice and assistance regarding recommended and new strategies to implement). In addition, if health care professionals are aware of the app, they can direct youth to the mobile phone app upon concussion diagnosis and during recovery. As a result, both end-user groups were involved in testing the usability of this app.

The scientific content of this app, including the concussion management strategies (eg, energy conservation, sleep, nutrition, and relaxation strategies) and supplementary concussion management information found within this prototype was adapted from the *Concussion & You* program. Through IID of the app prototype, the design team identified opportunities to add to the current *Concussion & You* content, for example, the design team found that a new concussion management strategies section could be a beneficial addition for concussed youth. As a result, the design team developed a beta concussion management strategies section titled *Social Goals*, which includes strategies based on the current best practices and expert opinion to help reduce the impact of social isolation and depression that may accompany a concussion diagnosis [6,38,39]. The user interface (UI) of this prototype was initially influenced by the *Concussion & You* program strategy planning tool found in the program's handbook [12], and this initial design was iteratively assessed by the design team using the IID approach, which resulted in the final UI that was used for the end-user usability study. For further information about the development process and design of the app see the dissertation *Development and Usability Testing of a Smartphone Technology for the Self-Management of Pediatric Concussion* [7].

The app's information architecture is shown in Figure 1 (see Multimedia Appendix 1 for a higher resolution image); the 8 main sections of the application are divided into 2 distinct groups: destinations that aid self-management actively through action (ie, Feelings, My Goals, Summary, and Set Reminder pages) and destinations for aiding self-management passively through information (ie, Concussion Library, Resources, Using NeuroCare, and Contact Experts). The key features of the app are that it guides concussed youth in creating a personalized concussion recovery self-management plan, allows youth to track how they are feeling each day, provides daily reminders, and provides feedback and recommendations on how youth can improve their concussion recovery. A key design principle required that navigation through the app should be intuitive, simple, and demand minimal engagement from the concussed youth to ensure the app could be used safely (ie, to avoid symptom exacerbation). For example, an alternative path is

indicated by the red arrows in Figure 1, which guides the user to complete the required daily actions using a short and concise workflow that requires minimal user engagement: these daily actions ask the user to select how they are feeling today on the *Feelings* page, go to the *My Goals* page, and select the competition status of each goal, and then navigate to the *Summary* page to check their progress. At the end of the week, youth are asked to review and revise their concussion recovery plan, which includes adding or removing goals based on guidance from their physician and the *Summary* page. The fully functional mobile phone prototype was further evaluated through this usability study with end users.

Figure 1. Final prototype information architecture. The app has a total of 8 main sections, with the Menu icons shown near the top. The arrows show how each screen is linked to the Menu, and how screens are linked to other screens within the application; the arrows indicate how a user could navigate through the different screens of the app.



Participants

A total of 14 participants were recruited for this study: 7 youth with a history of concussion and 7 health care professionals involved with concussion assessment, management, or research. Participants were excluded if they were younger than 10 years or older than 18 years; had not used mobile phone apps; were non-English speakers; were currently experiencing postconcussion symptoms; or if they had any physical, visual, or cognitive problems that may have precluded them from being able to use the mobile phone technology in the traditional way. Informed consent was obtained by all participants and/or their parents before participation. The study was approved by the Research Ethics Board at Holland Bloorview Kids Rehabilitation Hospital (REB#16-632) in Toronto, Canada.

Protocol

We conducted a formative lab-based usability study [32,40] with the fully functional mobile phone prototype. Formative lab-based usability testing is a widely used usability testing approach that is iterative in nature [32]; the goal of this testing is to make improvements in design before releasing the product [32]. This includes identifying and diagnosing the problems, making and implementing recommendations, and then re-evaluating the product [32]. In formative usability studies, the most significant usability findings are observed with the first 5 participants [41,42]. This study was conducted at either Holland Bloorview Kids Rehabilitation Hospital or at the health care professional's place of practice. The study was conducted in a quiet room and took 30 to 45 min to complete.

After the participant provided consent to participate in the study, they were asked to complete a demographics form. Then, the participant was introduced to the *think aloud* approach using a short video [43]. The *think aloud* approach asks the user to continuously verbalize their thoughts about their underlying thinking behind their interactions while using a technology by verbalizing what they are doing and why, stating when they encounter a problem, and how they feel while using the technology [43]. Next, the objective usability of mobile phone app was assessed by asking participants to complete 12 tasks (Table 1) using the mobile phone app while *thinking aloud*. During these tasks, the participants were audio-recorded, and a mobile phone screen recording app (AZ Screen Recorder [44]) was used to record the mobile phone screen; this included recording screen clicks and navigation to aid in identifying usability issues. The participant's actions as well as performance metrics, such as task success, the time on task, and the number of errors and assists were also observed and documented using pen and paper notetaking. In addition, any issues the participants faced while using the technology, including issue type, frequency, and severity were documented. After the completion of all the tasks, the subjective usability and usefulness of the users were assessed using the System Usability Scale (SUS) [45], a posttest questionnaire, and an exit interview (described below).

Table 1. Usability study tasks list. The usability issues or technology features that each task attempted to investigate are listed followed by the task instructions.

Task	Features tested
1	Using the visual scale: enter the application, and answer the "how are you feeling today" question with OKAY using the visual scale.
2	Finding and adding a goal on the *My Goals* page: add a specific *Social* goal to the action plan for a duration of 1 week.
3	Setting a reminder: set a reminder 1 min from the current time.
4	Responding to the reminder notification: respond to the notification reminder that was set in Task 3.
5	Using the visual scale and using the prompt (ie, toast notification): answer the "how are you feeling today?" question with GOOD. Then, use the prompt to go from Home screen to the *My Goals* page.
6	Setting goal competition status and using the prompt (ie, toast notification): set the completion status for goals on the *My Goals* page, and use the prompt to go from the *My Goals* page to the *Summary* page.
7	Finding and adding the recommended goal: add the physician-recommended goal on the *Summary* page to action plan for a duration of 2 days.
8	Deleting a goal from the action plan: delete the *Social* goal originally added in Task 2.
9	Setting goal completion status and using the prompt (ie, toast notification): set the completion status for goals on the *My Goals* page, and use the prompt to go from the *My Goals* page to the *Summary* page.
10a	Finding concussion education information: find the *concussion myths versus facts* educational page.
10b	Navigating to the Home screen: return to the Home screen of the app.
11	Reviewing and comprehending feeling and action plan history: find out how you were feeling yesterday, and identify the completion status of yesterday's goals.

Demographics Form

The demographics form for the 2 participant groups (youth and health care professionals) were customized for each group. Youths were asked questions regarding their concussion history and experience with managing a concussion, whereas health care professionals were asked about their involvement in concussion assessment, management, and/or research. For both participant groups, data on age, sex, and if they owned/had access to a mobile phone and/or tablet were collected. Furthermore, both groups were asked to answer questions about their perception of concussion knowledge and management in Canada using a 7-point Likert scale (1—strongly disagree, 7—strongly agree) and open-ended comments.

Usability and Usefulness

Usability is defined 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 [32,33]. The objective usability of the prototype was evaluated by measuring the extent to which the prototype could be used to complete specified tasks with effectiveness and efficiency [32,33], and the subjective usability was evaluated by measuring satisfaction using the SUS, posttest questionnaire, and exit interview. Subjective/perceived usefulness was measured using the posttest questionnaire and exit interview.

Objective Usability: Task Performance (Observational Notes, and Audio and Screen Recordings)

Task Success and Number of Errors

The effectiveness of this prototype was evaluated by measuring the number of errors and assists and the number of tasks that were completed successfully (task success). Task success is the most widely used usability performance metric; if a user cannot complete a given task, then there is likely a problem with the technology [32,33,40]. Errors and assists also indicate effectiveness; both are useful in pointing out particularly confusing parts of a technology [32,33,40,46]. Errors were defined as any action that caused the participant to deviate from the path to successful task completion. Assists were defined as any assistance provided to the participant to aid in task

completion; participants were provided assists only when they were having a considerable amount of difficulty with a task.

Time on Task

To measure the amount of effort (efficiency) with which participants completed each task, time on task was also measured, which included time for errors and corrections [32,33,40]; the faster the user completed a task, the lower the amount of effort required to complete a task, thereby offering an overall better experience. It is important to ensure that time is measured accurately and consistently [32,40,46]. We marked the start of each task as the time when the participant was told to start attempting the task, and the end time was marked as the time when the participant said "I am done"; waiting for the participant to say they are done is important, so that detectable usability issues do not go unidentified [32,40,46]. Although participants were asked to *think aloud* during this usability test, time-on-task data were still collected; however, using a concurrent *think aloud* protocol may impact task completion time. A solution is to ask participants to *hold* any longer comments until after a task is completed [32]; this solution was used to ensure task completion times were as accurate as possible while using the *think aloud* protocol. The number of, unnecessary actions or, actions exceeding the minimum number of actions required for a task is also indicative of the efficiency of a technology and were recorded [32,40,46]; it is possible for a task to have a fast completion time but still require a high amount of effort.

Subjective Usability and Usefulness: System Usability Scale, Posttest Questionnaire, and Exit Interview

Following the completion of the 12 tasks using the mobile phone app while *thinking aloud*, participants completed 2 posttest questionnaires, both focused on measuring the prototype's usability; the SUS [45] was issued first followed by a more general posttest questionnaire focused on usability and usefulness [32,47].

The SUS consists of 10 questions and uses a 5-point Likert-scale answering scheme to get a reliable and robust evaluation of a product [32,45,48]. The SUS is a validated and reliable measure of the subjective or perceived usability of a system with small sample sizes (ie, 8-12 users) [32,45,49]. The SUS questionnaire was modified to be more youth-friendly (age-appropriate language) and customized for assessing a mobile phone app.

Furthermore, an additional and more general posttest questionnaire was used to measure the usability and usefulness, that is, "the degree to which a product enables a user to achieve his or her goals, and is an assessment of the user's willingness to use the product at all" [50]. An adapted version of a usefulness questionnaire that was developed and validated by Davis was included as part of the posttest questionnaire to understand the perceived usefulness of this technology [50]. Additional questions were included to assess subjective satisfaction to complement the SUS score and better understand technology satisfaction. These questions were adapted from the Usefulness, Satisfaction, and Ease of Use questionnaire [32,51].

After completing the questionnaires, the participant was invited to take part in a semistructured exit interview; follow-up interviews are commonly used in usability studies where the researcher meets with the participants one-on-one to discuss in detail what the participant thinks about a specific topic in question, discuss usability issues, and clarify comments or behaviors exhibited during the usability study [43,52]. This exit interview was adapted from previous literature specific to usability evaluation [32,47,50,51,53], and it allowed the participant to share comments and opinions on the mobile phone technology, answer questions regarding the prototype's usability and usefulness, discuss any problems they encountered while completing the assigned tasks, assist in clarifying and resolving usability issues, and to clarify key comments or behaviors exhibited during the *think aloud* protocol.

For this study, the primary outcome measures were the task success, time on task, errors, and SUS scores. The task success and errors per task evaluated the effectiveness, whereas time on task measured efficiency, and satisfaction was evaluated using the SUS, posttest questionnaire, and the exit interview. The secondary outcome measures were the usability issues, unnecessary actions, assists, and the usefulness of the app.

Data Analysis

A triangulation approach [32,40,46] was used to identify the key usability issues with the mobile phone technology prototype. Descriptive statistics (eg, frequencies) were used to analyze all performance data (ie, task success, time on task, number of errors, unnecessary actions, and assists) and all close-ended demographic and posttest questionnaire data. Time-on-task data were analyzed using a measure of central tendency (ie, mean). Data from the think aloud protocol, exit interviews, screen recordings, and questionnaires for the tasks/questions indicating usability/usefulness issues were examined to identify the cause of, and the possible solutions for the issues using the approach described by Dumas and Redish [46]. The SUS questionnaire was analyzed using the procedure described by Brooke [45]; descriptive statistics (ie, measures of variability and central tendency) were also used to analyze the SUS scores and demographics data.

Results

Demographics

A total of 14 participants were recruited for this study: 7 youths with a history of concussion, and 7 health care professionals. Table 2 provides a summary of study participant demographics.

Table 2. Youth (n=7) and health care professional (n=7) demographics.

Participant group, category	Statistic
Youth	
Age (years), mean (SD)	12.7 (1.9)
Gender, female, n (%)	5 (71)
Months since most recent concussion, mean (SD)	18.9 (4.7)
Even with health care professionals helping me, I felt confused about what I should do to manage my concussion(s): agreed/strongly agreed, n (%)	5 (71)
I either own, or have daily access to, a smartphone, n (%)	7 (100)
I either own, or have daily access to, a tablet, n (%)	7 (100)
Health care professionals	
Age (years), mean (SD)	42.9 (15.7)
Gender, female, n (%)	7 (100)
Type of health care professional, n (%[a])	
Neuropsychologists	2 (29)
Occupational therapists	2 (29)
School nurses	2 (29)
Physical medicine and rehabilitation physician	1 (14)
Years of work experience in this role, mean (SD)	9.6 (7.3)
I find all youth in Canada are given enough information to manage their concussion, n (%)	
Disagreed/strongly disagreed	6 (86)
Slightly disagreed	1 (14)
I find pediatric concussion is managed in a consistent and standardized manner by all health care professionals in Canada: disagreed/strongly disagreed, n (%)	7 (100)
I either own, or have daily access to, a smartphone, n (%)	7 (100)
I either own, or have daily access to, a tablet, n (%)	4 (57)

[a]The total of the percentages sums to more than 100% due to rounding.

Objective Usability

Task Success

Mean task success rates were greater than 90% for 92% (11/12) of tasks, which indicates high usability. All participants successfully completed 7 of the 12 tasks (Table 1). A few participants were not able to successfully complete tasks 3, 4, 9, 10b, and 11. The percentage of participants who completed a task and their level of success (ie, zero problems [green], with 1 or more problems [blue], and task failure [red]) are shown in Figure 2.

Number of Errors

The frequency of assists (red), errors (orange), and actions (yellow) for each task are shown in Figure 3, which shows that a number of assists were provided to participants for task 4, and some assists were provided for task 10b and task 11, demonstrating severe usability issues with task 4, and moderate usability issues with tasks 10b and 11. One task failure occurred for task 3 (Figure 2), for which 0 assists were provided and only a few errors occurred. However, many actions were taken that exceeded the minimum number of actions required for task 3; this may point to a minor usability issue. One failure also occurred for task 9 (Figure 2), but there were 0 assists required, only 3 errors, and 5 unnecessary actions across 14 participants, further indicating that task 9 may be a minor usability issue. Tasks 5 and 7 had a 100% success rate but still exhibited some issues. Only 1 assist was provided for task 5, and a miniscule number of errors and unnecessary actions occurred. Therefore, this task is not likely to point to a usability issue, but the cause of the assist was still investigated. However, task 7 resulted in 9 errors, and many unnecessary actions were taken; task 7 may point to a usability issue.

Figure 2. Percentage of participants by levels of task success per task. Task 1: select "how are you feeling today" using the visual scale; Task 2: add a *Social* goal to the action plan; Task 3: set a reminder; Task 4: respond to the reminder; Task 5: select "how are you feeling today" using the visual scale, and use the toast notification to navigate; Task 6: set completion status for goals, and use the toast notification to navigate; Task 7: find and add the recommended goal; Task 8: delete a goal; Task 9: set completion status for all goals, and use the toast notification to navigate; Task 10: (a) find concussion education information and (b) navigate to the Home screen; and Task 11: review and comprehend feeling and action plan history.

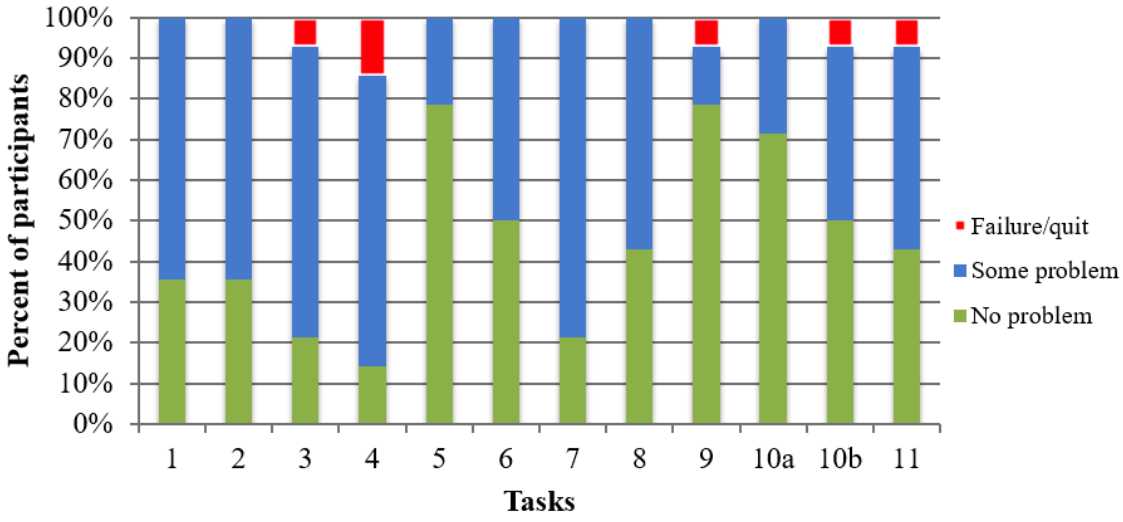
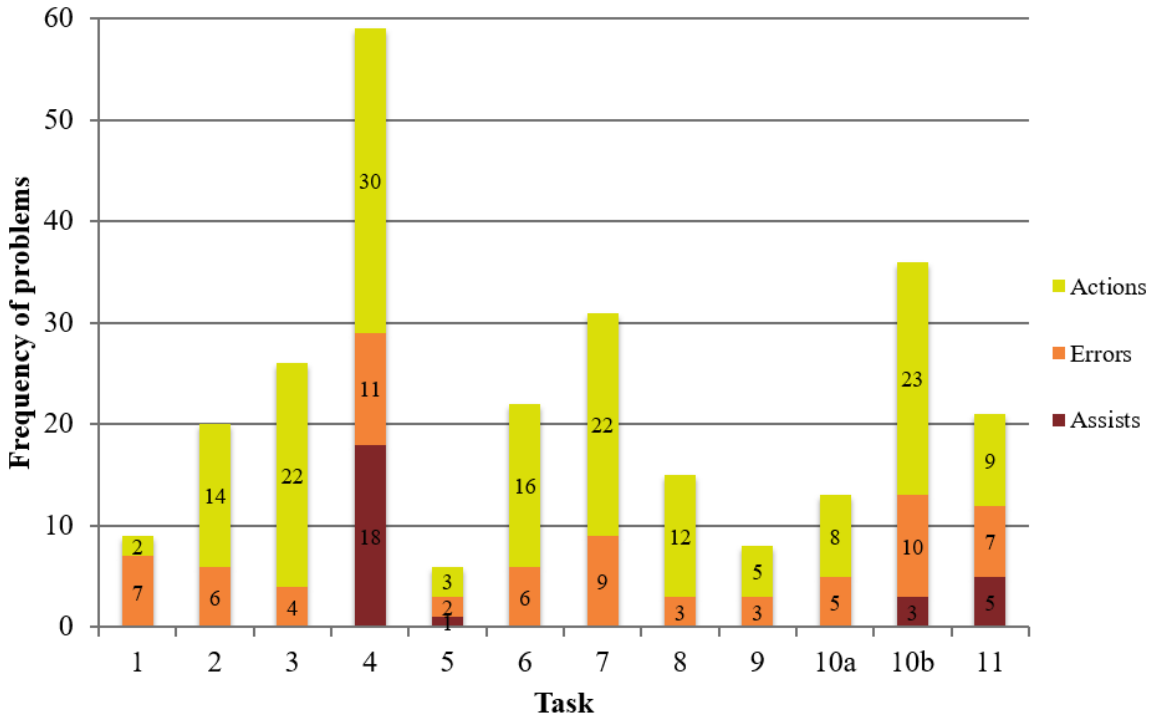


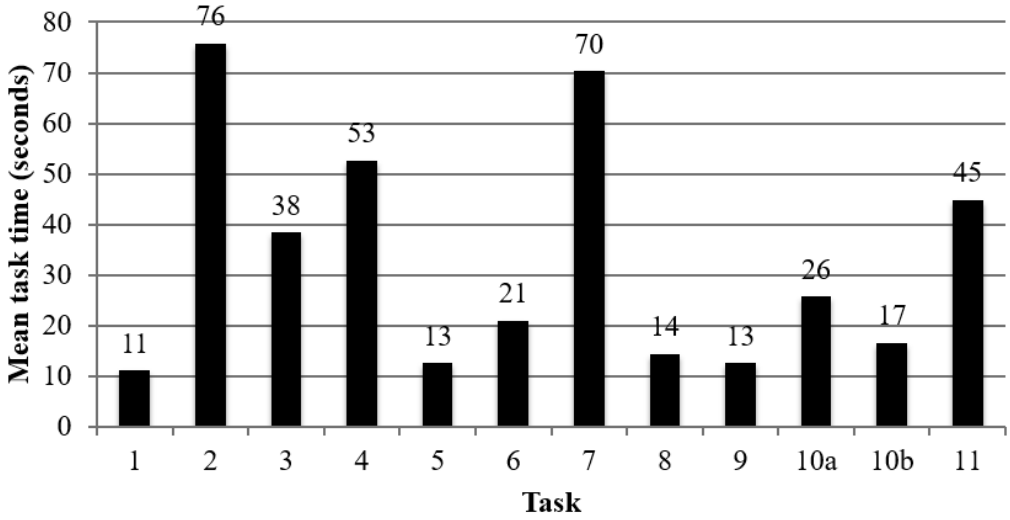
Figure 3. Frequency of assists, errors, and actions per task, across all 14 participants. Task 1: select "how are you feeling today" using the visual scale; Task 2: add a *Social* goal to the action plan; Task 3: set a reminder; Task 4: respond to the reminder; Task 5: select "how are you feeling today" using the visual scale, and use the toast notification to navigate; Task 6: set completion status for goals, and use the toast notification to navigate; Task 7: find and add the recommended goal; Task 8: delete a goal; Task 9: set completion status for all goals, and use the toast notification to navigate; Task 10: (a) find concussion education information and (b) navigate to the Home screen; and Task 11: review and comprehend feeling and action plan history.



Time on Task

The mean task completion time for each task is displayed in Figure 4. It was hypothesized that each task would take less than 2 min to complete; this hypothesis was confirmed. Figure 4 reveals that the mean time on task for each of the tasks was less than 77 seconds while using the *think aloud* protocol.

Figure 4. Mean task completion times per task, in seconds. Task 1: select "how are you feeling today" using the visual scale; Task 2: add a *Social* goal to the action plan; Task 3: set a reminder; Task 4: respond to the reminder; Task 5: select "how are you feeling today" using the visual scale, and use the toast notification to navigate; Task 6: set completion status for goals, and use the toast notification to navigate; Task 7: find and add the recommended goal; Task 8: delete a goal; Task 9: set completion status for all goals, and use the toast notification to navigate; Task 10: (a) find concussion education information and (b) navigate to the Home screen; and Task 11: review and comprehend feeling and action plan history.



Subjective Usability and Usefulness

System Usability Scale Questionnaire

Scores above 68 (SD 12.5) indicate above average usability [45,48,49]. The mean SUS score for this study was 81.9 (SD 11.3), indicating that, on average participants were highly satisfied with the usability of this mobile phone technology prototype. The mean SUS scores were calculated for the 2 groups, youth and health care professionals. The mean SUS score for youth was 87.5 (SD 8.5), whereas the mean SUS score for health care professionals was 76.4 (SD 11.5). SUS scores for 86% (6/7) of youth were equal to or above 82.5; 1 youth participant's SUS score was 72.5. However, SUS scores for only 43% (3/7) of health care professionals were equal to or above 82.5. Furthermore, 2 health care professionals had scores of 65 and 60, which is considered below average. A small but significant correlation between age and SUS scores showing that SUS scores decrease as age increases has been shown in the literature [54], which may partially explain the lower SUS scores among health care professionals.

Posttest Questionnaire

In the posttest questionnaire, youth participants (n=7) were asked that in the hypothetical case they experience another concussion in the future, if they would use this app; 71% (5/7) strongly agreed that they would use this app, and the remaining participants agreed (n=1) and slightly agreed (n=1). When youth and health care professionals were asked if they would recommend this application to a concussed youth, 100% (n=7) of youth either slightly agreed or agreed (6/7 agreed and 1/7 slightly agreed), and 4 health care professionals agreed that they would recommend this technology to concussed youth. However, 3 health care professionals did not agree that they would recommend this technology; 1 health care professional neither agreed nor disagreed (ie, neutral), 1 slightly disagreed, and 1 disagreed. To better understand these 3 ratings, the open-ended responses, if available, were reviewed. The health care professional who neither agreed nor disagreed stated that they did not know enough about the app to recommend it. The health care professional who slightly disagreed was concerned that using the technology (ie, screen time) and the amount of reading/cognition involved may exacerbate symptoms. The health care professional who disagreed also mentioned that if a youth was to have to choose between spending allowed screen time on this app versus school work, they would recommend the functional task over the use of this app. A concern among some (2/7) health care professionals who did not agree or slightly agreed to recommend this app was that recommending this app meant they were recommending screen time to concussed youth; this was associated with the fear that extended screen time could lead to exacerbation of symptoms. However, most (4/7) health care professionals stated that they would recommend this app to concussed youth.

Health care professionals and youths were asked if this technology would be useful in helping the youth self-manage their concussion: 86% of participants either slightly agreed or agreed (9/14 agreed and 3/14 slightly agreed) that this technology would be useful in helping the youth self-manage their concussion, 1 youth neither agreed nor disagreed, and 1 health care professional disagreed. Analyzing the open-ended answers from the questionnaire revealed that the youth who neither agreed nor disagreed thought it would be hard for youth to remember to set goals and change how they feel every day; however, reviewing the task success and errors data revealed that this youth failed to complete task 3 (ie, setting a daily reminder). The health care professional who disagreed provided no explanation for their choice to disagree with the usefulness of this app. However, during the exit interview, this health care professional did mention that they believed the technology would be useful in helping youth self-manage their concussions if the technology tracked postconcussion symptoms and somehow tied symptoms with the goals.

Discussion

Principal Findings

This research described the development and evaluation of a mobile phone app to aid youth in self-managing concussion. A fully functional mobile phone app prototype was developed, and a usability study was completed to evaluate this technology. Usability issues with this technology were identified, and actionable recommendations were provided to resolve the issues; these issues should be resolved to improve the usability of the technology. Furthermore, some overarching issues, and corresponding recommendations to further improve the app are discussed.

As discussed, some tasks led to task failure, requiring assists, errors, and/or unnecessary actions. The tasks were analyzed, beginning with the tasks that indicated high severity issues, followed by low to medium severity issues [32,40,46]. The recommendations to improve the design of the mobile phone technology were developed and listed for high severity issues (Table 3) and low to medium severity issues (Multimedia Appendix 2).

Table 3. High severity usability issues and recommendations.

Feature	Problem	Recommendation
Reminders	Many participants had difficulty finding the reminder because they attributed a reminder to something that would pop-up in the middle of the mobile phone's screen, emit a sound (*ringing*), and state that it is a reminder explicitly.	The reminder should explicitly state that it is the daily reminder that was set by the user from within the *NeuroCare* app, for example, the message within the reminder could state "Your daily reminder: How are you feeling today?"
Navigation menu	Some participants had difficulty locating/identifying the *main screen* of the app. Many participants attempted to click the *NeuroCare* brain logo in the navigation menu; participants thought that clicking this button should take them to the main screen. In addition, participants were attempting to look for a *home* icon to locate the main screen of the app.	The home screen of this technology is the "How Am I Feeling Today?" page. To ensure users can easily recognize that this is the main page, a *home* icon can be used to replace the current *smiley* icon for the "How Am I Feeling Today?" page. In addition, the brain logo in the navigation menu should be programmed such that when it is clicked, it takes the user to the main screen of the app.
Goals history	Participants were confused about the location of the goals history; the participants expected the goals history to be located on the *My Goals* page. However, when participants were asked if they thought it was useful to see their goals history in the *feelings calendar*, 100% (n=14) agreed.	The goals history should be moved to the *My Goals* page or the *Summary* page. Most participants expressed that they liked the calendar view, so it is recommended that the calendar format still be used to display the goal history.
Reminders (clock)	The current app clock-face was considered not intuitive and was too complex for participants, for example, some participants believed that the clock-face would only allow setting a reminder in 5-min intervals. Most participants mentioned that they identified more with a scrollable time picker and had difficulty using the clock-face time picker. One health care professional mentioned that the current clock-face might require a lot of cognitive effort.	To minimize the cognitive effort required to use this technology and improve usability, the current clock-face should be replaced with scrollable time picker.

In this study, the mean SUS score for this mobile phone app was found to be 81.9 (SD 11.3), which suggests that participants rated the app as highly usable; SUS scores above 68 (SD 12.5) indicate above average usability [45,48,49]. Sauro looked at the relationship between SUS scores and the *Net Promoter Score*. The latter asks individuals how likely they are to recommend a product to a friend or colleague [55]. Sauro found that individuals who rate a product with an SUS score of 82 (SD 5) tend to be *promoters* for the product [55]. Thus, the mean SUS score for this study of 81.9 suggests that people are likely to be *promoters* of this technology, and they are likely to recommend this technology to their friends or colleagues. More importantly, the high and consistent SUS scores provided by youth (mean 87.5, SD 8.5) suggest that they are more likely to be promoters of this technology than health care professionals (mean 76.4, SD 11.5). These results are in contrast to the SUS scores of a recently proposed intervention for pediatric concussion management titled SMART [24]. The SMART intervention was tested with 4 child/parent pairs, and the mean child SUS score was 81 (SD 22.8), whereas the mean parent score was 89 (SD 10.7) [24]. These scores suggest that the features and design of the SMART technology resonated better with older adults than children. In addition, the large SD in youth SUS scores indicates that some youth perceived the usability of the SMART technology as below average. In contrast, the results from our study suggest that all youth perceived the usability of the mobile phone technology as high and are likely to be promoters for this technology. Unlike the results for the SMART intervention, these results suggest that the features and design of this technology resonated better with youth than older adults. To better understand the discrepancy between SUS scores among health care professionals and youth, key responses to the posttest questionnaire were analyzed. In the posttest questionnaire, youth and health care professionals were asked if they would recommend this app to a concussed youth, 100% (n=7) of youth slightly agreed or agreed (6/7 agreed and 1/7 slightly agreed), and 4 health care professionals agreed that they would recommend this technology to concussed youth. However, 3 health care professionals did not agree that they would recommend this technology; 1 health care professional neither agreed nor disagreed (ie, neutral), 1 slightly disagreed, and 1 disagreed. These results support Sauro's claim that individuals with SUS scores of 82 (SD 5) tend to be *promoters* for the product and are more likely to recommend the product.

To better understand why some health care professionals did not completely agree to recommend the technology, the questionnaire's open-ended responses, if available, were reviewed. A concern among some health care professionals (2/7) who did not agree or slightly agreed to recommend this app was that recommending this app meant they were recommending screen time to concussed youth; this was associated with the fear that extended screen time could lead to exacerbation of symptoms. This may be due to the fact that the best practice concussion management guidelines recommend a period of physical and cognitive rest following a concussion [56]. However, new international consensus has suggested a shorter rest period; now, the suggested rest period is of approximately 24 to 48 hours after injury versus the previously suggested rest period until resolution of postconcussion symptoms [57]. In addition, the benefits of the rest period have not been validated [6], and it is unclear whether physical and cognitive rest aid concussed youth in recovery [58]. In addition, findings from recent studies suggest that prolonged rest after concussion is associated with increased risk for the development of secondary problems [58-60]; these secondary problems include, anxiety/stress, physical deconditioning, irritability, social isolation, and depression [6,61]. Furthermore, it is unknown the extent to which youth adhere to the recommendations for physical and cognitive rest [62]. However, further development of this mobile technology should aim to demand even lower cognitive effort to ensure the technology can be safely used by concussed youth. Many steps can be taken to reduce the amount of cognitive effort required for youth to use this technology, for example, resolving the identified usability issues can reduce the amount of time, frequency of errors, and the amount effort that is required to use a technology; the usability issues can be resolved by applying the provided recommendations (Table 3 and Multimedia Appendix 2). In addition, future iterations of the app should inform users about symptoms that may be exacerbated when using a mobile phone app (eg, screen time may lead to increased headaches, fatigue, light sensitivity, and difficulty concentrating), provide methods to reduce possibility of symptom exacerbation (eg, inform users to decrease screen brightness), and notify users of what actions they can take if using the app leads to symptom exacerbation (eg, in the result of symptom exacerbation, stop engaging with the app, rest and attempt to re-engage when symptom exacerbation has resolved). Nevertheless, this usability study instructed participants to complete a series of tasks sequentially, which could have led to a high perceived cognitive workload, whereas concussed youth would only be expected to complete a subset of these tasks every day. Concussed youth would be expected to complete tasks 1 and 5 (Table 1) everyday; these tasks ask youth to enter the app, select how they are feeling, then go to their *My Goals* page, and state the completion status of the goals in their action plan. In addition, the youth can view their *Summary* page, which is also a part of task 5. According to the task completion times (Figure 4), on average these 2 tasks together required 24 seconds to complete. At the end of every week, youth would be asked to perform task 2 or task 7: these tasks ask youth to add a new goal. On average, these tasks take approximately 70 seconds to complete for each goal. Thus, youth would be expected to use this technology for less than 1 min on a daily basis and less than 2 min at the end of each week; this suggests that this technology requires lower effort per day compared with other concussion management interventions [12,24]. For example, during the usability study for the SMART intervention, analyzing the time-on-module data revealed that a mean of 49 min was spent on completing 6 of the 8 modules (2 modules were missing timing data), and there were a total of 103 webpages across the 8 modules [24]. To evaluate how safely this technology can be used by concussed youth and reduce health care professionals' anxiety in recommending it, further work should include the analysis of perceived physical and cognitive workload as compared with other activities youth take part in during concussion recovery. A useful tool for evaluating perceived workload is the NASA Task Load Index questionnaire [63], which is a widely used and validated questionnaire [63] that can help to assess the perceived workload of this technology. In addition, further research is needed to trial this technology among a cohort of concussed youth to determine if the technology exacerbates postconcussion symptoms. Nevertheless, the majority (4/7) of health care professionals and all youth (7/7) stated that they would recommend this app to concussed youth.

In this study, health care professionals were asked about their perceptions of pediatric concussion management on the demographics form. When asked if they find all health care professionals in Canada manage pediatric concussion in a consistent manner, 100% (n=7) of the health care professionals disagreed or strongly disagreed with the statement. When health care professionals were asked if they find that all youth in Canada are given enough information to manage their concussion, 86% (6/7) of the health care professionals either disagreed or strongly disagreed with the statement; 1 health care professional slightly disagreed with the statement. These results are consistent with the current literature, which has shown that there is a lack of standardization and that significant gaps exist in the management of pediatric concussion in Canada [8-11]. The results from this study suggest that this technology may be useful in helping reduce the gaps in pediatric concussion management by providing easy access to expert-informed concussion management information and strategies and a tool that can guide youth in self-managing and tracking their concussion recovery.

All of the participants (n=14) in this study indicated that they either own or have daily access to mobile phone and tablet. This supports the findings from the recent Pew Internet & American Life Project that indicate that mobile phones have become the primary communication tool for the majority of adolescents in the United States [17,18]; 75% of those aged 12 to 17 years now own mobile phones [18]. Both youth and health care professionals have shown interest in this technology; as discussed earlier, 100% (n=7) of youth slightly agreed or agreed (6/7 agreed and 1/7 slightly agreed) and 4 health care professionals agreed that they would recommend this technology to concussed youth. In addition, 100% (n=7) of the youth participants agreed that they would use this technology if they were to suffer another concussion in the future, and 86% of participants either slightly agreed or agreed (9/14 agreed and 3/14 slightly agreed) that this technology would be useful in

helping the youth self-manage their concussion. This suggests that this mobile phone app may be an accessible, useful, and a feasible concussion management intervention for concussed youth.

This usability study provided valuable end-user feedback from both youth and health care professionals. A number of usability issues were identified, and the corresponding recommendations to improve the design of the app were provided (Table 3 and Multimedia Appendix 2); many low and moderate severity issues were identified, and 4 high severity issues were identified. In addition, some recommendations to improve the safety, uptake, and overall design of the technology were provided. Our findings suggest that participants rated this mobile phone app as having high subjective usability as indicated by a mean SUS score of 81.9 (SD 11.3). In addition, mean task success rates were greater than 90% for 92% (11/12) of tasks, and most (11/12; 92%) tasks had a total error frequency of less than 11, which also suggests high objective usability. On average, each task was completed in less than 2 min, which suggests this app is highly efficient. The results of the posttest questionnaires suggest that youth and health care professionals are open to using this app for self-management of concussion in youth and feel that this technology would be useful in helping the youth in managing their concussions. Overall, the results from the study suggest that participants rated this technology as usable, acceptable to users, and that it may be useful in helping youth self-manage concussion. Further work should include the analysis of perceived physical and cognitive workload to evaluate the safely of this technology, applying the recommendations to resolve the identified usability issues, modifying features to reduce physical and cognitive workload, and conducting a second usability study. In addition, further research is needed to trial this technology among a cohort of concussed youth to evaluate the effectiveness and safety of this technology.

Limitations

Most participants (12/14) were from Toronto, Ontario, and the participants who chose to take part in this study may have been more motivated, knowledgeable of concussion management, and comfortable with using mobile phones. Thus, this sample may not be representative of the general concussed youth and health care professional populations. We were unable to gain insight into how different age groups among children and youth may engage and rate usability differently as this would require a larger sample size with representation across ages. The data from usability study were analyzed and interpreted to identify usability issues. This could have biased the study results by not having interpreted a participant's comments appropriately [32,49]. However, to reduce this bias, we confirmed all findings during the exit interview and used the recordings to enhance and clarify the findings. This study was conducted with the researcher present in the room; a limitation of this type of study is that the behaviors and performance of participants may be altered as a result of their awareness of being observed [64]. Although the results of this study suggest that the participants' response to this mobile phone app has been very positive, further research is needed to trial this technology among a cohort of concussed youth to evaluate the effectiveness and safety of this technology and to identify the subpopulations for whom this intervention would be most effective.

Conclusions

This research describes the development and usability evaluation of an innovative and accessible pediatric concussion management intervention in the form of a fully functional Android mobile phone app prototype *NeuroCare*. The results from our usability study indicate that participants rated this technology as usable, acceptable to youth and health care professionals, and that it may be useful in helping youth self-manage concussion. Consistent with the current literature, results from this research suggest that there are large gaps in the way concussion is managed from both the youths' and health care professional's perspectives. This technology is expected to help bridge the gaps in pediatric concussion management by enabling and empowering youth to self-manage concussion by providing easy access to expert-informed concussion management strategies and helping guide youth in managing and tracking their concussion recovery. Next steps should include resolving the identified usability issues, modifying features to reduce cognitive and physical workload, and then conducting another usability study that should include the evaluation of perceived physical and cognitive workload. Future work should trial this technology among a cohort of concussed youth to determine the effectiveness and safety of this technology as a concussion self-management tool/intervention.

Acknowledgments

This work was funded by the Canadian Institutes of Health Research (CIHR; #127048), the Ontario Neurotrauma Foundation, and the Ontario Brain Institute. The Ontario Brain Institute is an independent nonprofit corporation, funded partially by the Ontario government. The opinions, results, and conclusions are those of the authors, and no endorsement by the Ontario Brain Institute is intended or should be inferred. We would also like to acknowledge the efforts of the members of the CIHR *NeuroCare* Team and the members of the Concussion Centre (Bloorview Research Institute, Bloorview, Ontario, Canada), specifically Dr Michelle Keightley and Talia Dick.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Final prototype information architecture (higher resolution image). The app has a total of 8 main sections, with the Menu icons shown near the top. The arrows show how each screen is linked to the Menu, and how screens are linked to other screens within the application; the arrows indicate how a user could navigate through the different screens of the app.

[PDF File (Adobe PDF File), 3MB - humanfactors_v6i2e12135_app1.pdf]

Multimedia Appendix 2

Low to medium severity usability issues and recommendations.

[PDF File (Adobe PDF File), 244KB - humanfactors_v6i2e12135_app2.pdf]

References

1. Zemek RL, Grool AM, Rodriguez DD, DeMatteo C, Rothman L, Benchimol EI, et al. Annual and seasonal trends in ambulatory visits for pediatric concussion in Ontario between 2003 and 2013. J Pediatr 2017 Feb;181:222-8.e2. [doi: 10.1016/j.jpeds.2016.10.067] [Medline: 27843008]
2. Fridman L, Scolnik M, Macpherson A, Rothman L, Guttmann A, Grool AM, et al. Annual trends in follow-up visits for pediatric concussion in emergency departments and physicians' offices. J Pediatr 2018 Jan;192:184-188. [doi: 10.1016/j.jpeds.2017.09.018] [Medline: 29150146]
3. Sahler CS, Greenwald BD. Traumatic brain injury in sports: a review. Rehabil Res Pract 2012;2012:659652 [FREE Full text] [doi: 10.1155/2012/659652] [Medline: 22848836]
4. Ontario Neurotrauma Foundation. 2010. Guidelines for Mild Traumatic Brain Injury and Persistent Symptoms URL: http://onf.org/system/attachments/60/original/Guidelines_for_Mild_Traumatic_Brain_Injury_and_Persistent_Symptoms.pdf [accessed 2019-04-08] [WebCite Cache ID 77THYlbY6]
5. Alexeev N. Brain Injury Professional. 2016. Models of Brain Injury Service Delivery from Around the Globe URL: https://issuu.com/bipmagazine/docs/bip39_internationall_service_delive [accessed 2019-04-08] [WebCite Cache ID 77THkkBcJ]
6. McCrory P, Meeuwisse W, Aubry M, Cantu B, Dvorak J, Echemendia RJ, et al. Consensus statement on concussion in sport—the 4th International Conference on Concussion in Sport held in Zurich, November 2012. Clin J Sport Med 2013 Mar;23(2):89-117. [doi: 10.1097/JSM.0b013e31828b67cf] [Medline: 23478784]
7. Sandhu H, Reed N, Mihailidis A. TSpace. Development and Usability Testing of a Smartphone Technology for the Self-Management of Pediatric Concussion 2017 Jan 1 URL: https://tspace.library.utoronto.ca/handle/1807/77895 [accessed 2019-05-21] [WebCite Cache ID 78XQ6fMAm]
8. Zemek R, Eady K, Moreau K, Farion KJ, Solomon B, Weiser M, et al. Knowledge of paediatric concussion among front-line primary care providers. Paediatr Child Health 2014 Nov;19(9):475-480 [FREE Full text] [Medline: 25414583]
9. Babul S. Addressing the need for standardized concussion care in Canada: Concussion Awareness Training Tool. Can Fam Physician 2015 Aug;61(8):660-662 [FREE Full text] [Medline: 26273074]
10. Stoller J, Carson J, Purcell L, Davidson B, Garel A, Bell M, et al. Do family physicans, emergency physicians and paediatricians give consistent sport-related concussion management advice? Br J Sports Med 2013 Mar 11;47(5):e1.16-e1.e1. [doi: 10.1136/bjsports-2012-092101.23]
11. Zonfrillo MR, Master CL, Grady MF, Winston FK, Callahan JM, Arbogast KB. Pediatric providers' self-reported knowledge, practices, and attitudes about concussion. Pediatrics 2012 Dec;130(6):1120-1125. [doi: 10.1542/peds.2012-1431] [Medline: 23147981]
12. Hunt AW, De Feo L, Macintyre J, Greenspoon D, Dick T, Mah K, et al. Development and feasibility of an evidence-informed self-management education program in pediatric concussion rehabilitation. BMC Health Serv Res 2016 Dec 17;16(1):400 [FREE Full text] [doi: 10.1186/s12913-016-1664-3] [Medline: 27534848]
13. Spanakis EG, Santana S, Tsiknakis M, Marias K, Sakkalis V, Teixeira A, et al. Technology-based innovations to foster personalized healthy lifestyles and well-being: a targeted review. J Med Internet Res 2016 Jun 24;18(6):e128 [FREE Full text] [doi: 10.2196/jmir.4863] [Medline: 27342137]
14. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
15. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
16. Proudfoot J. The future is in our hands: the role of mobile phones in the prevention and management of mental disorders. Aust N Z J Psychiatry 2013 Feb;47(2):111-113. [doi: 10.1177/0004867412471441] [Medline: 23382507]
17. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]

18. Lenhart A, Ling R, Campbell S, Purcell K. Pew Research Center. 2010 Apr 20. Teens and Mobile Phones | Pew Research Center's Internet & American Life Project URL: https://www.pewinternet.org/2010/04/20/teens-and-mobile-phones/ [WebCite Cache ID 77TIgTjGc]
19. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. J Med Internet Res 2003 Dec 18;5(4):e32 [FREE Full text] [doi: 10.2196/jmir.5.4.e32] [Medline: 14713660]
20. Worthen-Chaudhari L, McGonigal J, Logan K, Bockbrader MA, Yeates KO, Mysiw WJ. Reducing concussion symptoms among teenage youth: evaluation of a mobile health app. Brain Inj 2017;31(10):1279-1286 [FREE Full text] [doi: 10.1080/02699052.2017.1332388] [Medline: 28665690]
21. Babcock L, Byczkowski T, Wade SL, Ho M, Mookerjee S, Bazarian JJ. Predicting postconcussion syndrome after mild traumatic brain injury in children and adolescents who present to the emergency department. JAMA Pediatr 2013 Feb;167(2):156-161 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.434] [Medline: 23247384]
22. Babcock L, Kurowski BG, Zhang N, Dexheimer JW, Dyas J, Wade SL. Adolescents with mild traumatic brain injury get SMART: an analysis of a novel web-based intervention. Telemed J E Health 2017 Dec;23(7):600-607 [FREE Full text] [doi: 10.1089/tmj.2016.0215] [Medline: 28112591]
23. Kurowski BG, Wade SL, Dexheimer JW, Dyas J, Zhang N, Babcock L. Feasibility and potential benefits of a web-based intervention delivered acutely after mild traumatic brain injury in adolescents: a pilot study. J Head Trauma Rehabil 2016;31(6):369-378 [FREE Full text] [doi: 10.1097/HTR.0000000000000180] [Medline: 26360000]
24. Dexheimer JW, Kurowski BG, Anders SH, McClanahan N, Wade SL, Babcock L. Usability evaluation of the SMART application for youth with mTBI. Int J Med Inform 2017 Dec;97:163-170 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.007] [Medline: 27919376]
25. Zasler N, Katz DI, Zafonte RD. Brain Injury Medicine, Second Edition: Principles And Practice. New York: Demos Medical; 2012.
26. Jonassaint CR, Shah N, Jonassaint J, De Castro L. Usability and feasibility of an mHealth intervention for monitoring and managing pain symptoms in sickle cell disease: the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]
27. Astegic. 2013 Apr 12. Native vs HTML5 – looked at objectively, the debate is over URL: https://www.astegic.com/native-vs-html5-looked-objectively-debate/ [accessed 2019-04-08] [WebCite Cache ID 77TIVkWPR]
28. Maguire M. Methods to support human-centred design. Int J Hum-Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
29. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
30. Caburnay CA, Graff K, Harris JK, McQueen A, Smith M, Fairchild M, et al. Evaluating diabetes mobile applications for health literate designs and functionality, 2014. Prev Chronic Dis 2015;12:E61 [FREE Full text] [doi: 10.5888/pcd12.140433] [Medline: 25950568]
31. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. J Biomed Inform 2016 Feb;59:115-129 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.008] [Medline: 26639894]
32. Albert W, Tullis T. Measuring The User Experience: Collecting, Analyzing, And Presenting Usability Metrics (Interactive Technologies). San Diego, CA: Morgan Kaufmann; Jul 29, 2013.
33. Law E, Bevan N, Christou G, Springett M, Lárusdóttir M. Classifying and Selecting UX and Usability Measures. 2008 Presented at: International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM); June 18, 2008; Reykjavik, Iceland p. 13-18.
34. Nielsen J. Nielsen Norman Group. 2012. Usability 101: Introduction to Usability URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2019-04-08] [WebCite Cache ID 77TIuZaya]
35. Microsoft Corporation. Usability in Software Design URL: https://docs.microsoft.com/en-us/previous-versions/ms997577(v=msdn.10) [accessed 2019-04-08] [WebCite Cache ID 77TIziE2K]
36. Larman C. Agile And Iterative Development: A Manager's Guide. Boston: Addison-wesley Professional; Aug 21, 2003.
37. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 1995 May 1;152(9):1423-1433 [FREE Full text] [Medline: 7728691]
38. Daneshvar DH, Riley DO, Nowinski CJ, McKee AC, Stern RA, Cantu RC. Long-term consequences: effects on normal development profile after concussion. Phys Med Rehabil Clin N Am 2011 Nov;22(4):683-700, ix [FREE Full text] [doi: 10.1016/j.pmr.2011.08.009] [Medline: 22050943]
39. Chrisman SP, Richardson LP. Prevalence of diagnosed depression in adolescents with history of concussion. J Adolesc Health 2014 May;54(5):582-586 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.10.006] [Medline: 24355628]
40. Rubin J, Chisnell D. Handbook Of Usability Testing: How To Plan, Design, And Conduct Effective Tests, Second Edition. Indianapolis, IN: Wiley; 2008.
41. Nielsen J, Mack R. Usability Inspection Methods. New York: John Wiley & Sons; 1994.

42. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems.: ACM; 1993 May 1 Presented at: CHI'93; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213. [doi: 10.1145/169059.169166]
43. Nielsen J. Usability Engineering. San Francisco, CA: Morgan Kaufmann; 1994.
44. Google. AZ Screen Recorder URL: https://play.google.com/store/apps/details?id=com.hecorat.screenrecorder.free&hl=en [accessed 2019-04-08] [WebCite Cache ID 77VrSYMjn]
45. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, USA: CRC Press; 1995:189-194.
46. Dumas J, Redish J. A Practical Guide to Usability Testing. England: Intellect Ltd; 1999.
47. Usability.gov. 2016. Planning a Usability Test URL: https://www.usability.gov/how-to-and-tools/methods/planning-usability-testing.html [accessed 2019-04-08] [WebCite Cache ID 77TK7ZiV0]
48. Brooke J. SUS: a retrospective. J Usability Stud 2013 Feb;8(2):29-40 [FREE Full text]
49. Tullis T, Stetson J. A Comparison of Questionnaires for Assessing Website Usability. 2004 Presented at: Usability Professionals Association (UPA) 2004; June 7-11, 2004; Minneapolis, Minnesota URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.396.3677&rep=rep1&type=pdf
50. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
51. Lund A. ResearchGate. 2001. Measuring Usability with the USE Questionnaire URL: https://www.researchgate.net/publication/230786746_Measuring_Usability_with_the_USE_Questionnaire [accessed 2019-04-08] [WebCite Cache ID 77TKgXAu4]
52. Rohrer C. Nielsen Norman Group. 2014. When to Use Which User-Experience Research Methods URL: https://www.nngroup.com/articles/which-ux-research-methods/ [accessed 2019-04-08] [WebCite Cache ID 77TKY7E6F]
53. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Design, development, and formative evaluation of a smartphone application for recording and monitoring physical activity levels: the 10,000 Steps. Health Educ Behav 2013 Apr;40(2):140-151. [doi: 10.1177/1090198112449460] [Medline: 22984196]
54. 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 [FREE Full text]
55. Sauro J. A Practical Guide To The System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, California: Createspace Independent Publishing Platform; 2011.
56. Hingley S, Ross J. Guidelines for diagnosing and managing paediatric concussion: Ontario Neurotrauma Foundation guideline. Arch Dis Child Educ Pract Ed 2016 Apr;101(2):58-60. [doi: 10.1136/archdischild-2014-307252] [Medline: 26297033]
57. McCrory P, Meeuwisse W, Dvorak J, Aubry M, Bailes J, Broglio S, et al. Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med 2017 Apr 26;51(11):838-847. [doi: 10.1136/bjsports-2017-097699]
58. Thomas DG, Apps JN, Hoffmann RG, McCrea M, Hammeke T. Benefits of strict rest after acute concussion: a randomized controlled trial. Pediatrics 2015 Feb;135(2):213-223 [FREE Full text] [doi: 10.1542/peds.2014-0966] [Medline: 25560444]
59. Kirkwood MW, Yeates KO, Wilson PE. Pediatric sport-related concussion: a review of the clinical management of an oft-neglected population. Pediatrics 2006 Apr;117(4):1359-1371. [doi: 10.1542/peds.2005-0994] [Medline: 16585334]
60. Schneider KJ, Iverson GL, Emery CA, McCrory P, Herring SA, Meeuwisse WH. The effects of rest and treatment following sport-related concussion: a systematic review of the literature. Br J Sports Med 2013 Apr;47(5):304-307. [doi: 10.1136/bjsports-2013-092190] [Medline: 23479489]
61. Kuehl MD, Snyder AR, Erickson SE, McLeod TC. Impact of prior concussions on health-related quality of life in collegiate athletes. Clin J Sport Med 2010 Mar;20(2):86-91. [doi: 10.1097/JSM.0b013e3181cf4534] [Medline: 20215889]
62. Giza CC, Hovda DA. The new neurometabolic cascade of concussion. Neurosurgery 2014 Oct;75(Suppl 4):S24-S33 [FREE Full text] [doi: 10.1227/NEU.0000000000000505] [Medline: 25232881]
63. Hart S, Staveland L. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in Psychology, Human Mental Workload. New York: Elsevier Science & Technology; 1988:139-183.
64. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Abbreviations

CIHR: Canadian Institutes of Health Research
IID: iterative and incremental development
mHealth: mobile health
SUS: System Usability Scale
UI: user interface

Edited by G Eysenbach; submitted 20.09.18; peer-reviewed by K Deldar, J Alvarez Pitti, S Wade; comments to author 07.01.19; revised version received 09.04.19; accepted 15.04.19; published 31.05.19.

Please cite as:
Sandhu H, Wilson K, Reed N, Mihailidis A
A Mobile Phone App for the Self-Management of Pediatric Concussion: Development and Usability Testing
JMIR Hum Factors 2019;6(2):e12135
URL: http://humanfactors.jmir.org/2019/2/e12135/
doi:10.2196/12135
PMID:31152527

Original Paper

Development of a Clinical Interface for a Novel Newborn Resuscitation Device: Human Factors Approach to Understanding Cognitive User Requirements

Laura Pickup[1*], MSc, PhD; Alexandra Lang[2*], BSc, PhD; Lara Shipley[3*], MBChB (Hons), MRCS, MRCPCH; Caroline Henry[3*], B Med Sci (Hons), BMBS (Hons), MRCPCH; James Carpenter[4], MEng, PhD; Damon McCartney[4], PhD; Matthew Butler[4], PhD; Barrie Hayes-Gill[5], BSc, PhD, FIET, CEng, Eurlng; Don Sharkey[3], B Med Sci (Hons), BMMS, FRCPCH, PhD

[1]LP Human Factors Ltd, Taunton, United Kingdom

[2]Trent Simulation and Clinical Skills Centre, Nottingham Universities Hospitals NHS Trust, Queen's Medical Centre, Nottingham, United Kingdom

[3]Division of Child Health, Obstetrics and Gynaecology, School of Medicine & Health Sciences, University of Nottingham, Nottingham, United Kingdom

[4]SurePulse Medical Limited, Medicity, Nottingham, United Kingdom

[5]Optics and Photonics Research Group, Department of Electrical and Electronic Engineering, University of Nottingham, Nottingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Barrie Hayes-Gill, BSc, PhD, FIET, CEng, Eurlng
Optics and Photonics Research Group
Department of Electrical and Electronic Engineering
University of Nottingham
University Park
Nottingham, NG7 2RD
United Kingdom
Phone: 01159515547
Email: Barrie.Hayes-Gill@nottingham.ac.uk

Abstract

Background: A novel medical device has been developed to address an unmet need of standardizing and facilitating heart rate recording during neonatal resuscitation. In a time-critical emergency resuscitation, where failure can mean death of an infant, it is vital that clinicians are provided with information in a timely, precise, and clear manner to capacitate appropriate decision making. This new technology provides a hands-free, wireless heart rate monitoring solution that easily fits the clinical pathway and procedure for neonatal resuscitation.

Objective: This study aimed to understand the requirements of the interface design for a new device by using a human factors approach. This approach combined a traditional user-centered design approach with an applied cognitive task analysis to understand the tasks involved, the cognitive requirements, and the potential for error during a neonatal resuscitation scenario.

Methods: Fourteen clinical staff were involved in producing the final design requirements. Two pediatric doctors supported the development of a visual representation of the activities associated with neonatal resuscitation. This design was used to develop a scenario-based workshop. Two workshops were carried out in parallel and involved three pediatric doctors, three neonatal nurses, two advance neonatal practitioners, and four midwives. Both groups came together at the end to reflect on the findings from the separate sessions.

Results: The outputs of this study have provided a comprehensive description of information requirements during neonatal resuscitation and enabled product developers to understand the preferred requirements of the user interface design for the device. The study raised three key areas for the designers to consider, which had not previously been highlighted: (1) interface layout and information priority, as heart rate should be central and occupy two-thirds of the screen; (2) size and portability, to enable positioning of the product local to the baby's head and allow visibility from all angles; and (3) auditory feedback, to support visual information on heart rate rhythm and reliability of the trace with an early alert for intervention while avoiding parental distress.

Conclusions: This study demonstrates the application of human factors and the applied cognitive task analysis method, which identified previously unidentified user requirements. This methodology provides a useful approach to aid development of the clinical interface for medical devices.

(JMIR Hum Factors 2019;6(2):e12055) doi:10.2196/12055

KEYWORDS

neonatal resuscitation; medical device; human factors; user-centred design; applied cognitive task analysis

Introduction

Background

Globally, there are approximately 3.6 million neonatal deaths annually (ie, in the first 28 days of life), with 70% occurring on the first day of life [1]. Up to 10% of newborns (79,000/year in the United Kingdom and 13 million/year worldwide) require some form of resuscitation at birth, with an estimated 7 million babies worldwide requiring more advanced resuscitation [2]. The correct structured management of resuscitation in the first few "golden" minutes after birth is critical to prevent significant morbidity (eg, cerebral palsy due to hypoxia) or death. There is strong evidence that standardized resuscitation training and algorithms significantly improve newborn outcomes and could reduce mortality by up to 30% [2,3].

International newborn resuscitation guidelines highlight the importance of using the heart rate (HR) to guide resuscitation and stabilization methods [4]. However, many of the methods used to measure HR are inaccurate or technically challenging, particularly in premature infants.

When assessing the HR, practitioners always have access to a stethoscope but use other technologies less frequently [5]. HR assessment using the stethoscope, through auscultation, is inaccurate in about one-third of cases [6,7] and is not continuous; therefore, it needs to be performed every 30 seconds. As such, it is time consuming, which pauses resuscitation and can lead to errors.

This paper describes an enquiry investigating the design and use of a novel medical device developed to address the unmet need of standardizing and facilitating neonatal resuscitation. In an emergency time-critical resuscitation situation where failure can mean death of an infant, it is vital that clinicians are provided information in a timely, precise, and clear manner to support decision making. The nature of this context requires an interface that can ensure both the efficiency and reliability of staff to access the most critical information. A touchscreen interface was considered to be the best hardware solution. The work described here focuses on the development of the design requirements for a touchscreen interface that is integral to this novel medical device.

To understand the requirements of this new device, and specifically, the contributors to interface design, a human factors approach was implemented, which combined a traditional user-centered design approach with an applied cognitive task analysis (ACTA) [8]. The aim of this study was to understand not only the tasks involved but also the cognitive requirements of clinicians. This study has enabled the generation of an interface specification. In addition, the study's findings provide points of learning to other medical device developers and clinicians, with an aim of understanding the complex requirements and information needs of clinicians during neonatal resuscitation.

Medical Devices to Measure Heart Rate in Neonates

Other common techniques for monitoring HR in the neonatal intensive care unit, such as electrocardiography or pulse oximetry, were not developed for resuscitation at birth. These systems are used less due to their reliability, delay in HR readings, and practical issues (eg, difficulty ensuring adhesion to the skin) [5,9].

In the delivery room, electrocardiography and pulse oximetry sensors are connected to the main monitors by cables. This can make attachment more challenging and risks cold exposure with the potential for hypothermia, which is an independent risk factor for death in premature babies [10]. Current resuscitation guidance for premature babies highlights the prevention of hypothermia, and therefore, priority is given to drying the baby's head, putting on a hat, and placing the body (wet) in a plastic bag/wrap [11-13].

To address the issues described, a novel HR monitoring hardware solution using reflectance mode optical photoplethysmograpy, an optical sensor, has been designed. This monitor has been integrated into a single-use newborn hat, specifically for use in newborn babies requiring resuscitation. This solution aims to fit naturally into the existing care pathway, allowing wireless, hands-free, quick, continuous, and accurate HR monitoring via a touchscreen interface as well as minimizing the risk of hypothermia. The effectiveness of the solution is a combination of two features: the forehead placement, where blood flow is preserved even in babies with a low HR (the forehead blood supply comes from the carotid arteries that supply the brain), and the sensor's patented optical arrangement and signal processing scheme, which has been proven to provide high signal quality from neonatal patients [14]. Additionally, the hat uses wireless communication, allowing greater flexibility in deployment than cable-based solutions.

Human Factors/Ergonomics in Medical Device Design

The value of human factors/ergonomics (HFE) integration to medical device design and patient safety has been recognized over recent years [15-19], has gained formal recognition in standards [20-23], and is a requirement of the European Medical Devices Directive 93/42 and its 2007 amendment for obtaining Conformité Européenne approval. Concerns still remain about the quality and effectiveness of the interpretation of all relevant standards and integration of HFE within the design/development process. There appears to be a lack of "exemplar case studies"

to illustrate how the design process and user-centered design can contribute to product design in health care and how HFE should be routinely implemented [24-26]. This is acknowledged with specific barriers identified within small and medium enterprises such as university spin-out companies [18].

This study contributes to the body of evidence on the application of HFE methods to the formative evaluation of this novel medical device, as required by the relevant standards [23], through a collaboration between a university spin-out company, the School of Medicine, and the Human Factors Research group in the University of Nottingham.

This study focuses specifically on understanding human-computer interactions and user requirements for the computer interface of the device. The aims were (1) to identify gaps in existing knowledge on user requirements for the interface design of a novel resuscitation device and (2) to represent the key design requirements to promote usability of the touchscreen interface of the device.

Methods

This study collected and analyzed data from intended and representative future users of the new device.

The ACTA method was selected for this study, as it is known to be beneficial to health care domains [8,26]. The ACTA facilitates the elicitation of cognitive requirements from clinicians relative to a particular task and translates them into design requirements for system designers [8]. The ACTA has four key stages. Table 1 highlights how each stage of the method is relevant to understanding the task of neonatal resuscitation. For the purposes of this study, the ACTA method was modified to accommodate clinical working practices, and the simulation interview took the form of an interactive scenario-based workshop.

The workshop aimed to recruit a range of health care professionals, with varying levels of experience and representative of those who might have involvement in neonatal resuscitation procedures. A convenience sampling approach was adopted for the recruitment of participants from two large tertiary-based teaching hospitals in the United Kingdom. Posters and flyers advertised the details of the study, and 12 staff with experience of neonatal resuscitation were successfully recruited (Table 2).

To explore the cognitive requirements further and elicit insight from all practitioners, the workshop protocol divided the practitioners into two groups of six people, split evenly to ensure equal numbers of each job role for each group with the exception of the clinical educator and trainee who were put in different groups. This allowed different levels of experience and job roles to explore the same simulation (Textbox 1). Ethical approval was provided by the University of Nottingham, and all participants gave their informed consent.

The two researchers (LP and AL) familiarized themselves with the task of neonatal resuscitation by observing videos of a simulated resuscitation provided by the two subject matter experts (SMEs; LS and CH who are neonatal doctors with 8 years of resuscitation experience) and follow-up interviews to clarify points of uncertainty and task identification. This was necessary for practical reasons, as the observation of such events cannot be planned. A review of the national neonatal resuscitation algorithm [11] provided the researchers with an understanding of the current UK practice. Finally, relevant international and British standards [20-23] were consulted to provide direction for the designers on medical device recommendations.

There were five outputs from this study that were achieved through the products listed in Table 1:

1. A high-level representation of the tasks required to identify the need and completion of neonatal resuscitation (Table 1 - Stage 1 - task diagram
2. Identification of the key/difficult cognitive requirements for neonatal resuscitation tasks, critical information, and decision points (Table 1 - task diagram and knowledge audit interview)
3. Analysis of the cognitive demand associated with key tasks and potential errors (Table 1 - knowledge audit and simulation interview)
4. User opinion on interface design options to support cognitive requirements, reduce potential for error, and record neonatal resuscitation events (Table 1 - simulation interview)
5. A comprehensive outline of user and design requirements for the interface design and relevant standards (Table 1 - simulation interview and cognitive demands table)

Table 1. Description of the applied cognitive task analysis.

Description	Stage 1 - Task diagram	Stage 2 - Knowledge audit interview	Stage 3 - Simulation interview - workshop	Stage 4 - Cognitive demands table
Method	Interviews with two SMEs[a] familiar with the task of neonatal resuscitation • Interview 1: Task identification with SME 1 (150 minutes) • Interview 2: Verification of task representation with SME 2 and identification of key/difficult cognitive tasks with SMEs 1 and 2 (75 minutes)	Interview with 2 SMEs (180 minutes). Starting with the use of the knowledge audit probes (Multimedia Appendix 1) to elicit general domain knowledge of how an expert may deal with a neonatal resuscitation while exploring potential errors that novice users may make. Specific examples of how certain cues and strategies supported individual tasks were also explored.	Observation of a challenging scenario (Textbox 1) involving the task of neonatal resuscitation. Each key task is queried to explore the critical cues, assessment, actions, and potential for error: • What actions, if any, would you take at this point? • What do you think is going on here? What is your assessment of the situation at this point in time? • What pieces of information led you to this situation assessment and these actions? • What errors would an inexperienced person be likely to make in this situation?	To summarize and integrate the information obtained from the previous three steps and interview data gathered prior to the study.
Purpose	To provide a broad view of the task and identify difficult cognitive components	To highlight which aspects of the task require expertise and which cues and strategies are relied upon to understand the impact on the novice user	To determine the cognitive process involved with key tasks and potential error	A comprehensive record of the findings of the project goals
Products	Key tasks associated with neonatal resuscitation using sticky notes (Figure 1) Visual representation tasks using Microsoft Visio software (version 2013) Verification of task representation Key/difficult cognitive tasks (eg, those requiring decision making, judgements, assessments, or problem solving)	Identification of critical cues and interpretation of information to diagnose and predict situation Identification of strategies relied upon by expert users Identification of the potential for errors in novice users	Identification of difficult cognitive components of task, information, and priorities Identification of critical cues relevant to decision making for each key task and potential for errors in novice users Essential and desirable information requirements Group mock-up of interface design on a cardboard model Individual annotation of a paper-based image of the intended interface screen	A spreadsheet of the data collated through the study

[a]SME: subject matter expert.

Table 2. Details of the workshop participants.

Job role	Number of years/range of experience	Number of participants
Neonatal trainee nurse	2	1
Neonatal nurse	1-16	2
Midwife	0-25	4
Pediatric/neonatal doctor	1.5-5	2
Neonatal clinical nurse educator	30	1
Advanced neonatal nurse practitioner	15-22	2

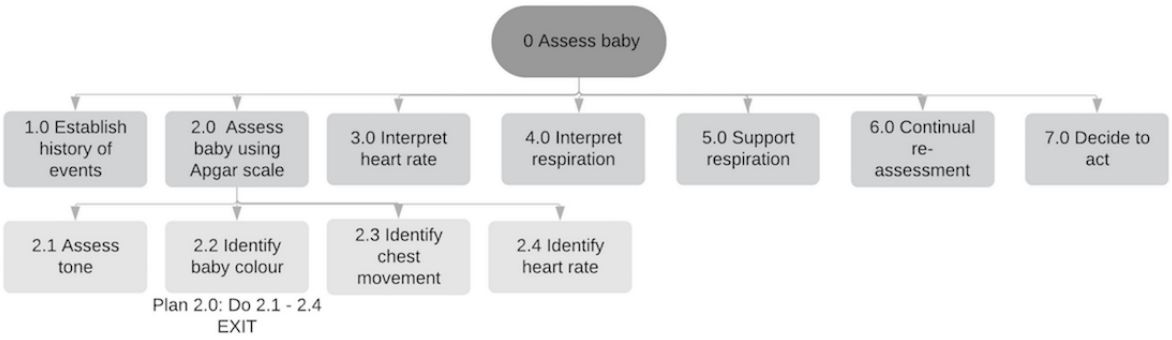
Textbox 1. Simulation of a challenging scenario used to probe practitioners during the workshop.

Past Clinical History

A first-time mother at 42 weeks' gestation presents with her baby stuck due to shoulder dystocia. She has a slight fever and no past medical history but has received diamorphine during her labor. Labor was induced through artificial rupture of the membrane. She has prolonged rupture of membranes and labor has been ongoing for 24 hours. The baby's head was delivered 10 minutes before the baby's body and the airway appears clear.

1. **Assessment of observations**
   - Baby presents floppy, white, and not responding to vigorous stimulus. Heart rate < 60 beats/minute with stethoscope and no respiratory effort evident.
2. **Progression of intervention**
   - After two sets of five inflation breaths, there is still no chest movement.
3. **Chest compressions commenced**
   - The chest moved but heart rate remained slow.

Figure 1. A sample of the representation of the task analysis completed with clinicians to illustrate neonatal resuscitation.



Results

The practitioner workshop involved pediatric doctors (n=2), neonatal nurses and educators (n=3), advanced neonatal nurse practitioners (n=2), and midwives (n=4) with 0-30 years of experience (average of 11 years). The workshop and SME interviews identified factors relevant to device and interface design not previously considered by the design team.

Context of Neonatal Resuscitation Tasks

The critical characteristics of a neonatal resuscitation were described by practitioners as time pressured and unpredictable albeit well-rehearsed. This is likely to be an emotional and stressful situation for the parents involved. The location of a neonatal resuscitation may vary and could include the labor suite, midwife-led unit, birthing pool room, operating theatre, patient's home, or ambulance. Participants suggested that the portability of the system should therefore be prioritized. Practitioners considered the attachment features of the device to replicate those of a car satellite navigation screen, with options to secure (which implies the device is attached but can be adjusted) and rotate the screen to ensure a continual good line of sight. The physical environment suggests lighting may also vary (eg, bright theatre lights and variability in lighting within a single resuscitation). The device may also remain in use when transferring the baby between delivery and the neonatal unit or in ambulances, hence making it resistant to vibration and movement.

Alarms were considered useful only in certain contexts and the type, such as frequency, and pitch of alarm required sensitivity in design to avoid undesirable consequences for parents and clinicians [27,28].

The device may be used in the context of other medical devices (eg, Resuscitaire, a portable "platform" for neonatal resuscitation integrated with required equipment). Compatibility between equipment is essential to ensure usability and reliability.

Tasks Relevant to Neonatal Resuscitation

The SME interviews (interviews 1 and 2 in Table 1) suggested seven relevant high-level cognitive tasks and produced the first study output, a high-level representation of the tasks required during neonatal resuscitation (Multimedia Appendix 2): establish a history of events, assess baby, interpret HR, interpret respiration, support respiration, continual reassessment, and decide to act.

These tasks were considered based on the core principles of the task analysis [29]. The top layer represents "what" has to be done (Multimedia Appendix 2), and further descriptions in the layer below describe "how" it has to be done (Multimedia Appendix 2). A visual representation was shared with clinicians

within the workshops, and a consensus was reached for the final presentation. Figure 1 illustrates a sample of the representation shared.

Nine tasks in total were agreed upon by the SMEs and workshop participants, to have a cognitive element to them (Multimedia Appendix 2). The nine tasks included receive antenatal history, assess baby using Apgar score [30], interpret chest movement, interpret HR, decide on action, direct view and clearing of airway, assist breathing, decide to intubate, and decide to medicate.

Cognitive Requirements, Demands, and Potential Error

The interviews completed during the development of the task diagram and the knowledge audit interview elicited information relevant to the difficulty and nature of the cognitive work including cues, assessment, judgements, problem solving, decisions, and actions combined with potential challenges and errors and strategies relevant to the nine cognitive tasks.

The findings from these first two stages of the ACTA method were verified and enriched by data obtained during the simulation interview workshop. The data from all three stages of the ACTA method were collated within a spreadsheet (a template of the one used is provided in Multimedia Appendix 2) and then combined and simplified to produce a cognitive demand table for each of the nine cognitive tasks (Textbox 2). These created the second and third study outputs.

Textbox 2. Cognitive demand table to assess the baby using the Apgar score to inform decision and actions.

Why is it difficult?

- Interpretation of heart rate and chest movement relative to normal parameters
- Judgement of accuracy and reliability of heart rate display
- Reliance on previous experience and recognition of “normal” heart rate and chest movement to inform decision and actions
- Multiple tasks in short time frame: visual check of heart rate, chest movement, tone, and skin color. Continual re-evaluation every 30 seconds
- Requires expertise to ensure decision making within a short timeframe and potentially stressful environment

Common errors

- Accuracy in interpretation of heart rate
- Fail to recall normal heart rate and chest movement
- Estimation/recall of time elapsed between key events
- Failure to recognize when to act (eg, call for help and intubation)
- Avoidance behavior: fear to act/“failure to rescue”
- Overreliance on technology (lacking reliability) and colleague’s earlier assessment
- Quiet breathing missed (eg, preterm babies)
- Lighting can distort baby color

Cues

- Absent heart rate, heart rate < 60 beats/minute
- Heart rate > 100 beats/minute
- Floppy
- White coloring
- No breathing/gasping
- Stressful environment
- Absence of baby crying

Strategies

- Consider how to obtain support with minimal alarm to parent
- Continually question interpretation/reliability of information
- Continual reassessment at 30 seconds and after 1, 5, and 10 minutes
- Closer inspection (eg, ear to mouth), observe rib cage and abdomen, listen for absence of sound or gasping

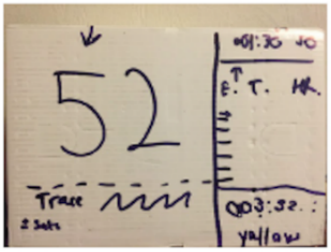
The two workshop groups further informed the fourth and final study outputs, which produced a specification and illustrated mock ups with notable differences in the priorities for design. One group preferred to protect the simplicity of the device and represent HR as the only physiological marker, with an event timeline running in the background. The other group preferred to include oxygen saturation and a visible record of an event timeline (Figure 2).

After the two workshop groups had worked through the simulation independently, they together presented their specifications and mock ups. Individual participants were then asked to reflect on the work they had done in their groups and on the presentation from the other groups to produce their own personal interface design. These individual contributions were analyzed to interpret group preferences and produce cumulative representations of the data as a heat map (Figure 3). This indicates consensus on the location of interface information sources, summarizing individual location preferences (12 practitioners) of the five information types. The x-axis indicates the width of the screen and the y-axis indicates the height of the screen. Each screen was broken up into 24 areas (6 along the x-axis and 4 along the y-axis). The color bars are normalized against the maximum number of practitioner votes for each area on the heat map.

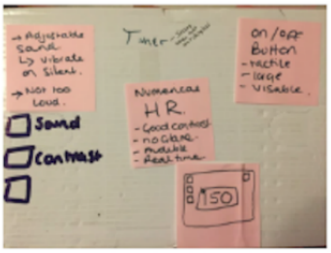
In summary, the stages within the ACTA method are provided below:

1. Task diagram
   - A breakdown of the physical and cognitive activities involved in the context of neonatal resuscitation
   - Visual representation of SME perspectives of the key cognitive activities
2. Knowledge audit interviews
   - Detailed descriptions of the nature of the cognitive work required, cues, and strategies relied upon and potential for error
   - Insight into differences between expert and novice practitioners in the context of neonatal resuscitation
   - Examples of previous experiences that revealed influences of the people involved/present, environmental factors, and the emotional nature of the context.
3. Simulation interview (workshop)
   - Verification of the task diagram and understanding of cognitive requirements based on a broader group of experts
   - Additional insights into cognitive work required, cues and strategies relied upon, and potential for error based on the experience of the participants
   - Consideration of future user and health care contexts
   - Design suggestions that reflect practitioners' preferences and understanding of the cognitive activities and potential errors discussed
4. Cognitive demands table
   - An assimilation of all of the abovementioned findings in a usable format to inform and justify the development of the design requirements

Figure 2. Mock-up of interface designs produced during simulation interview.



Practitioner group A mock-up



Practitioner group B mock-up

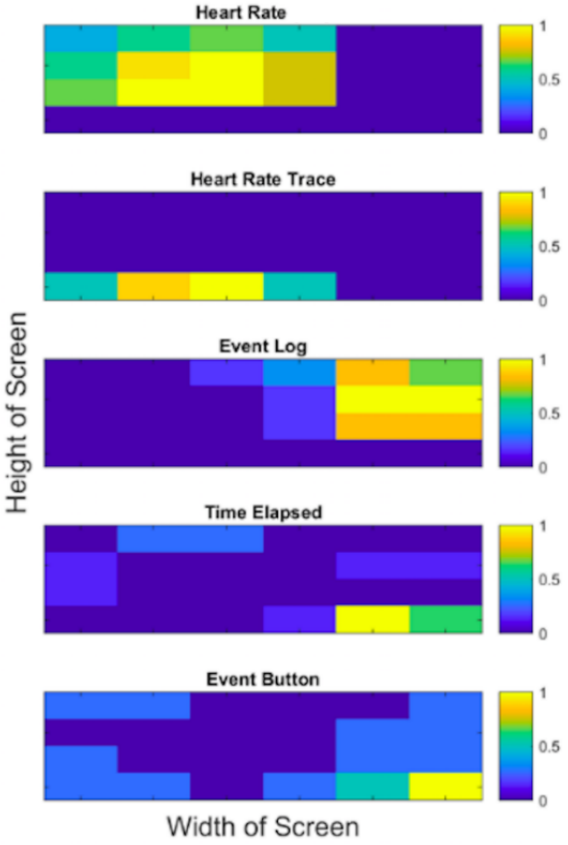
Workshop Design Interface mock up A

(numbers are arbitrary)

Workshop Design Interface mock up B

(numbers are arbitrary)

Figure 3. Heat map indicating consensus for the location of key information sources.



Themes relative to essential and desirable characteristics for the interface were elicited. These were combined with recommendations from international standards to produce a set of design requirements [20,23,31-35]. A typical example of the information contained within these requirements is provided in Textbox 3. The final decision for timer position was influenced by users and optimization of the display screen space. The information obtained informed the final design developed and indicated priorities for future usability testing. This information and the heat map were developed as a block diagram and informed the graphical user interface concept (Figure 4).

Textbox 3. Design requirements related to information layout.

Essential characteristics

Heart rate should be central and the largest text on the screen. This should allow visibility from all angles and the heart rate information should occupy two-thirds of the screen.

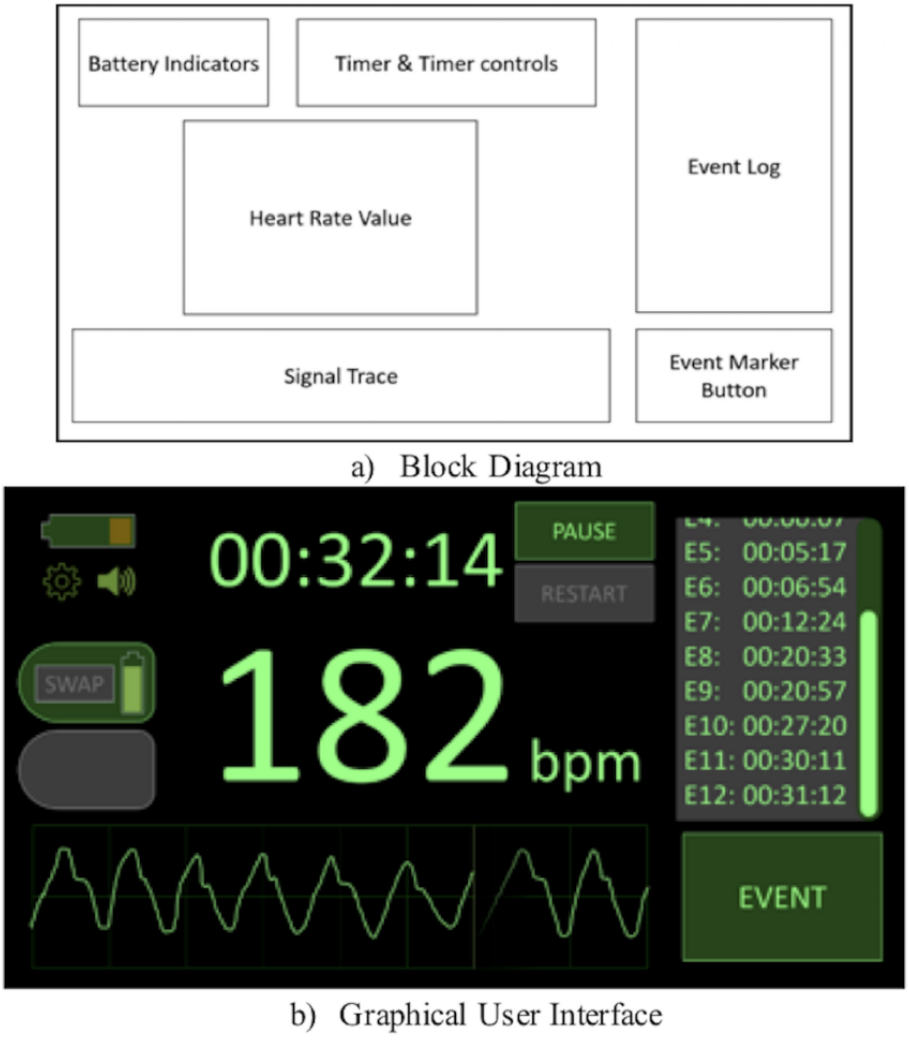
Desirable characteristics

Divide the screen to have a margin on the side of the screen with buttons to mark events. One group suggested illustration of a visible timeline. The second group did not agree with any more information than essentially required (eg, heart rate).

International standards and recommendations

Hierarchy of the content of information displayed should be implied by the layout. The most important information should occupy priority space, typically top left for large screens and central for smaller screens, with adequate blank space and borders to separate information sources

Figure 4. Block diagram (a) and the Graphical User Interface concept (b) developed from the block diagram.



Discussion

Principal Findings

To our knowledge, the ACTA method has not been used in the development of resuscitation devices. There are many HFE methods relevant to the design process [36,37]. The value of methods suitable for identifying user requirements was considered previously [17], and that paper concluded that both focus groups and user testing were beneficial. ACTA in medical device design does not appear to be well applied [18] despite recognition that it could be a useful tool in the domain of health care [8]. The application of such methods, rather than just the completion of the traditional hierarchical task analysis, is well recognized for their benefit of formatively understanding critical cues, decision making, judgements, constraints, and potential errors in the context of a work situation; however, they are also considered resource intensive [8,38,36].

The ACTA was developed to address some of these issues. The method was developed to allow practitioners within the area of work studied and system designers to elicit cognitive requirements relative to task performance and translate these into design requirements [8].

This approach was considered desirable for this study as a method to understand user decision making and critical information requirements from the intended interface and to illustrate a method that could be applied by practitioners themselves in future design/evaluation of medical devices. In addition, pragmatically, the time available suggested that efficiency was desirable in any method selected, which the ACTA offered.

This method has allowed the tasks required for neonatal resuscitation to be fully considered in relation to cognitive requirements, actions, and potential errors. This participatory approach has offered a systematic analysis of the resuscitation process, described as “logical and rigorous” by the SMEs. Successful implementation within the context of health care has been suggested as benefiting from such participation to ensure that relevant stakeholders influence the design of an intervention to fit their own contexts [39]. ACTA allowed user requirements to be identified specific to different contexts and stages within the resuscitation process.

The success of the ACTA approach came from engaging participants from different job roles to consider contexts familiar to them and ensure practitioners considered cues with the greatest significance to completing the required tasks, likely

errors, and how interface design can support these tasks. The outputs of this study have provided a comprehensive description of information requirements during neonatal resuscitation and enabled product developers to understand the core and preferred requirements of the user interface design for the device. These outputs have been used to develop an interface, which prioritizes simplicity and provides a set of user requirements, to test the device during future testing (Figure 5).

The study raised three key areas for the designers to consider, which had not previously been identified: interface layout and information priority, size and portability of the device, and auditory feedback.

The amount of information, which ultimately influences the size of the screen, will be determined by the intended function of the technology. Considering the task of neonatal resuscitation, it becomes apparent that early on in the process of resuscitation, HR is the key indicator used by practitioners. This information was prioritized by both groups and all individual designs of the interface. Some preferred that this alone should be the function of the device. It was considered desirable to ensure the device had a relatively small interface that could be positioned freely and local to the baby's head. The auditory feedback proposed by practitioners was to support visual information and interventions early on in the resuscitation process. The nature of the feedback should communicate information on HR, such as rhythm and reliability of the trace. The practitioners went on to suggest that with different auditory settings, the device could be used as a monitoring device within a neonatal unit, not previously considered by the developers. The implications of auditory feedback raised the importance of considering both practitioners and patient representatives (eg, parents within future usability testing) [39].

An integral timer was also considered essential, as it would serve as an indicator for the timeline of events within a resuscitation. Currently, the clock started at the time of birth is integral to the Resuscitaire, but future user testing needs to explore how the novel device will influence this practice.

Considering the community setting, there may be less access to oxygen saturation devices. The novel HR device has the potential to compensate for this absence. Within secondary care, the oxygen saturation devices were considered useful when intubating, to secure an airway for the transfer to a neonatal unit. Including oxygen saturations implied greater usability for the device in alternative situations and work contexts in the future. However, practitioners also acknowledged a risk where oxygen saturations could actually be a distraction for less experienced practitioners from the critical cue of HR, which is considered a better indicator of neonatal status and relied upon by expert practitioners.

The benefit of having an HR reading immediately next to the baby's head on the wireless module was considered of high value. This would reduce the need to continually look between the baby's head and an interface screen, where the HR would be viewed at a distance. Suggestions were made about the functionality of the wireless module (Figure 1), including how it could be the component within the device used to download contextual information such as an event log. This would avoid the need for a separate memory stick or disc to store data, reducing the risk of lost device components or information, which is a current problem with other medical devices.

Failure to recognize or acknowledge a deteriorating HR was considered. The device design could incorporate feedback to increase awareness of this critical cue. A change in screen color was suggested as the preferred prompt by some; however, further usability testing is required to find out if this improves practitioner performance in reacting to a deteriorating HR or is perceived as distracting or anxiety provoking. How color is used on the interface will also need to be explored by using color convention guidance and usability testing [20,23,32,33].

Figure 5. Resuscitation device comprises a single-use thermally insulating hat that communicates wirelessly (via battery powered modules) with a display mounted on the resuscitation table.

The final decisions made on the interface layout were based on the optimization of the screen by the design team. "Future proofing" the device was also considered during the workshop and generated an enthusiastic discussion on how an additional, but linked, mobile device could be used to assist a designated scribe in recording key events within the resuscitation timeline. Currently, the accuracy of a written report of neonatal resuscitation is variable, usually involving the most inexperienced team member to scribe. Future electronic records were proposed as complementary to an event marker control, operated by those delivering or supporting the resuscitation, while recognizing that the timing may be slightly delayed. However, this was considered sufficient to develop a retrospective view of the sequence of events. The electronic recording of these data was considered of significant value for those in governance roles, clinical learning, and audit.

Strengths and Limitations

The ACTA method provides an efficient, comprehensive, and participatory approach capable of understanding user decision making and critical information requirements from the intended interface. Practitioners discussed the potential for this device beyond the original context considered by the developers.

The limitations of the study were in the sampling of clinicians. Volunteers led to the larger ratio of nurses and midwives to doctors. However, the SMEs were experienced doctors in neonatal resuscitation and fully engaged in the whole study. Limitations of this study have constrained further development of the interface, and device, simulation, and usability testing should ensure the views and suggestions raised by participants can be tested and translated into a real-world device.

Conclusions

This is the first study to apply the ACTA approach to elicit user requirements for a novel device for neonatal resuscitation. This study demonstrates the application of human factors to inform the development of resuscitation devices, and more generally, for medical device developers and clinicians in the design and evaluation of medical technologies.

The study has provided previously unidentified user requirements and details about the variables, which will inform future usability testing of the interface developed.

Acknowledgments

This work was supported by the Medical Research Council (Grant number: MRC MC_PC_ 15012) and the Engineering and Physical Sciences Research Council (Grant number: EP/K 503101/1) via a Tioga Ltd PhD CASE award. This work was additionally supported by the Faculty of Engineering, University of Nottingham's Hermes Fellowship scheme, and the National Institute for Health Research (NIHR), MindTech Healthcare Technology Cooperative. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the MRC, the NIHR, or the Department of Health. The authors would like to thank all the participants and collaborators of this project. LP and AL were involved in the design and execution of the study. LS and CH were integral to the data collection and analysis of data. JC and DM contributed to the background and images relevant to the medical device. All authors contributed to the review and writing of the paper.

Conflicts of Interest

JC reports on other work by SurePulse Medical Limited; outside the submitted work, he is the CEO of this company. BRHG and DS are unpaid nonexecutive directors of SurePulse Medical Limited, and JC, BRHG, and DS are shareholders. DM and MB are employees of Surepulse Medical Ltd. SurePulse Medical Limited are the developers of vital sign monitoring technologies for newborns from the University of Nottingham. The company Surepulse Medical Limited has a filed an international patent for this device (WO 2017/149325).

Multimedia Appendix 1

Knowledge audit probe.

[PDF File (Adobe PDF File), 66KB - humanfactors_v6i2e12055_app1.pdf]

Multimedia Appendix 2

Cognitive tabular task analysis.

[XLS File (Microsoft Excel File), 97KB - humanfactors_v6i2e12055_app2.xls]

References

1. Lawn JE, Kerber K, Enweronu-Laryea C, Cousens S. 3.6 million neonatal deaths--what is progressing and what is not? Semin Perinatol 2010 Dec;34(6):371-386 [FREE Full text] [doi: 10.1053/j.semperi.2010.09.011] [Medline: 21094412]
2. Lee ACC, Cousens S, Wall SN, Niermeyer S, Darmstadt GL, Carlo WA, et al. Neonatal resuscitation and immediate newborn assessment and stimulation for the prevention of neonatal deaths: a systematic review, meta-analysis and Delphi

estimation of mortality effect. BMC Public Health 2011 Apr 13;11 Suppl 3:S12 [FREE Full text] [doi: 10.1186/1471-2458-11-S3-S12] [Medline: 21501429]
3. Xu T, Wang H, Ye H, Yu R, Huang X, Wang D, et al. Impact of a nationwide training program for neonatal resuscitation in China. Chin Med J (Engl) 2012 Apr;125(8):1448-1456. [Medline: 22613652]
4. Wyckoff MH, Aziz K, Escobedo MB, Kapadia VS, Kattwinkel J, Perlman JM, et al. Part 13: Neonatal Resuscitation: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015 Nov 03;132(18 Suppl 2):S543-S560. [doi: 10.1161/CIR.0000000000000267] [Medline: 26473001]
5. Mann C, Ward C, Grubb M, Hayes-Gill B, Crowe J, Marlow N, et al. Marked variation in newborn resuscitation practice: a national survey in the UK. Resuscitation 2012 May;83(5):607-611 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.01.002] [Medline: 22245743]
6. Voogdt KGJA, Morrison AC, Wood FE, van Elburg RM, Wyllie JP. A randomised, simulated study assessing auscultation of heart rate at birth. Resuscitation 2010 Aug;81(8):1000-1003 [FREE Full text] [doi: 10.1016/j.resuscitation.2010.03.021] [Medline: 20483522]
7. Chitkara R, Rajani AK, Oehlert JW, Lee HC, Epi MS, Halamek LP. The accuracy of human senses in the detection of neonatal heart rate during standardized simulated resuscitation: implications for delivery of care, training and technology design. Resuscitation 2013 Mar;84(3):369-372 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.07.035] [Medline: 22925993]
8. Militello LG, Hutton RJ. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. Ergonomics 1998 Nov;41(11):1618-1641. [doi: 10.1080/001401398186108] [Medline: 9819578]
9. Kamlin COF, Dawson JA, O'Donnell CPF, Morley CJ, Donath SM, Sekhon J, et al. Accuracy of pulse oximetry measurement of heart rate of newborn infants in the delivery room. J Pediatr 2008 Jun;152(6):756-760 [FREE Full text] [doi: 10.1016/j.jpeds.2008.01.002] [Medline: 18492509]
10. Cantwell R, Clutton-Brock T, Cooper G, Dawson A, Drife J, Garrod D, et al. Saving Mothers' Lives: Reviewing maternal deaths to make motherhood safer: 2006-2008. The Eighth Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. BJOG 2011 Mar;118 Suppl 1:1-203 [FREE Full text] [doi: 10.1111/j.1471-0528.2010.02847.x] [Medline: 21356004]
11. Resuscitation Council (UK) 2015. URL: https://www.resus.org.uk/resuscitation-guidelines/ [accessed 2018-08-24]
12. Patel D, Piotrowski ZH, Nelson MR, Sabich R. Effect of a statewide neonatal resuscitation training program on Apgar scores among high-risk neonates in Illinois. Pediatrics 2001 Apr;107(4):648-655. [Medline: 11335738]
13. Saugstad OD, Ramji S, Rootwelt T, Vento M. Response to resuscitation of the newborn: early prognostic variables. Acta Paediatr 2005 Jul;94(7):890-895 [FREE Full text] [doi: 10.1111/j.1651-2227.2005.tb02007.x] [Medline: 16188811]
14. Grubb MR, Carpenter J, Crowe JA, Teoh J, Marlow N, Ward C, et al. Forehead reflectance photoplethysmography to monitor heart rate: preliminary results from neonatal patients. Physiol Meas 2014 May;35(5):881-893. [doi: 10.1088/0967-3334/35/5/881] [Medline: 24742972]
15. Lin L, Isla R, Doniz K, Harkness H, Vicente KJ, Doyle DJ. Applying human factors to the design of medical equipment: patient-controlled analgesia. J Clin Monit Comput 1998 May;14(4):253-263. [Medline: 9754614]
16. Lin L, Vicente KJ, Doyle D. Patient Safety, Potential Adverse Drug Events, and Medical Device Design: A Human Factors Engineering Approach. Journal of Biomedical Informatics 2001 Aug;34(4):274-284. [doi: 10.1006/jbin.2001.1028] [Medline: 11977809]
17. Garmer K, Ylvén J, MariAnne Karlsson I. User participation in requirements elicitation comparing focus group interviews and usability tests for eliciting usability requirements for medical equipment: a case study. International Journal of Industrial Ergonomics 2004 Feb;33(2):85-98. [doi: 10.1016/j.ergon.2003.07.005]
18. Martin JL, Barnett J. Integrating the results of user research into medical device development: insights from a case study. BMC Med Inform Decis Mak 2012 Jul 19;12:74 [FREE Full text] [doi: 10.1186/1472-6947-12-74] [Medline: 22812565]
19. Gurses AP, Ozok AA, Pronovost PJ. Time to accelerate integration of human factors and ergonomics in patient safety. BMJ Qual Saf 2012 Apr;21(4):347-351. [doi: 10.1136/bmjqs-2011-000421] [Medline: 22129929]
20. International Organization for Standardization. 2007. ISO 14971:2007 Medical devices – Application of risk management to medical devices URL: https://www.iso.org/standard/38193.html [accessed 2019-05-24]
21. Gov.uk. 2017. Human Factors and Usability Engineering Guidance for Medical Devices Including Drug-Device Combination Products URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/645862/HumanFactors_Medical-Devices_v1.0.pdf [accessed 2019-06-05]
22. FDA. 2017. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/media/80481/download [accessed 2019-05-24]
23. AAMI. 2015. ANSI/AAMI/ IEC 62366- 1:2015 Medical devices – Part 1: Application of usability engineering URL: https://my.aami.org/aamiresources/previewfiles/6236601_1503_preview.pdf [accessed 2019-05-24]
24. Privitera MB, Evans M, Southee D. Human factors in the design of medical devices - Approaches to meeting international standards in the European Union and USA. Appl Ergon 2017 Mar;59(Pt A):251-263. [doi: 10.1016/j.apergo.2016.08.034] [Medline: 27890135]

25. Buckle P, Clarkson PJ, Coleman R, Ward J, Anderson J. Patient safety, systems design and ergonomics. Appl Ergon 2006 Jul;37(4):491-500. [doi: 10.1016/j.apergo.2006.04.016] [Medline: 16753132]
26. Martin JL, Norris BJ, Murphy E, Crowe JA. Medical device development: the challenge for ergonomics. Appl Ergon 2008 May;39(3):271-283. [doi: 10.1016/j.apergo.2007.10.002] [Medline: 18061139]
27. Edworthy J. Medical audible alarms: a review. J Am Med Inform Assoc 2013 May 01;20(3):584-589 [FREE Full text] [doi: 10.1136/amiajnl-2012-001061] [Medline: 23100127]
28. Edworthy J, Hellier E. Fewer but better auditory alarms will improve patient safety. Qual Saf Health Care 2005 Jun;14(3):212-215 [FREE Full text] [doi: 10.1136/qshc.2004.013052] [Medline: 15933320]
29. Kirwan BL. A Guide To Task Analysis: The Task Analysis Working Group. London: CRC Press; 1992.
30. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg 1953;32(4):260-267. [Medline: 13083014]
31. Ergonomics of human-system Interaction Part 210: Human-centred design for interactive systems. 2010. URL: https://www.sis.se/api/document/preview/912053/ [accessed 2019-05-24]
32. Ergonomic design of control centres Part 5: Displays and controls. URL: https://www.sis.se/api/document/preview/909962/ [accessed 2019-05-24]
33. Ergonomic principles in the design of work systems. URL: https://www.sis.se/api/document/preview/920898/ [accessed 2019-05-24]
34. AAMI. 2006. General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems URL: https://my.aami.org/aamiresources/previewfiles/606010108_1310_preview.pdf [accessed 2019-05-24]
35. Association for the Advancement of Medical Instrumentation. 2009. ANSI/AAMI HE75:2009, Human factors engineering – Design of medical devices URL: http://s3.amazonaws.com/rdcms-aami/files/production/public/FileDownloads/Products/HE750910_TOC.pdf [accessed 2019-05-24]
36. Stanton N. Human Factors Methods. Surrey: Routledge; 2013.
37. Wilson JN, Corlett E. Evaluation of Human Work: A practical ergonomics methodology. London: CRC Press; 1995.
38. Salmon P, Jenkins D, Stanton N, Walker G. Hierarchical task analysis vs. cognitive work analysis: comparison of theory, methodology and contribution to system design. Theoretical Issues in Ergonomics Science 2010 Nov;11(6):504-531. [doi: 10.1080/14639220903165169]
39. Goodyear-Smith F, Jackson C, Greenhalgh T. Co-design and implementation research: challenges and solutions for ethics committees. BMC Med Ethics 2015 Nov 16;16:78 [FREE Full text] [doi: 10.1186/s12910-015-0072-2] [Medline: 26573410]

Abbreviations

ACTA: applied cognitive task analysis
HFE: Human Factors/Ergonomics
HR: heart rate
SME: subject matter expert

Edited by A Kushniruk; submitted 30.08.18; peer-reviewed by D Furniss, T Henriques; comments to author 10.11.18; revised version received 14.01.19; accepted 22.01.19; published 08.06.19.

Please cite as:
Pickup L, Lang A, Shipley L, Henry C, Carpenter J, McCartney D, Butler M, Hayes-Gill B, Sharkey D
Development of a Clinical Interface for a Novel Newborn Resuscitation Device: Human Factors Approach to Understanding Cognitive User Requirements
JMIR Hum Factors 2019;6(2):e12055
URL: http://humanfactors.jmir.org/2019/2/e12055/
doi:10.2196/12055
PMID:31199321

Original Paper

A Hazard Analysis of Class I Recalls of Infusion Pumps

Xuemei Gao[1], ME; Qiang Wen[2], ME; Xiaolian Duan[3], MD; Wei Jin[1], ME; Xiaohong Tang[2], BA; Ling Zhong[2], ME; Shitao Xia[2], BE; Hailing Feng[1], BSc; Daidi Zhong[1], PhD

[1]Bioengineering College, Chongqing University, Chongqing, China

[2]Department of Medical Device Adverse Event Vigilance, Chongqing Center for Adverse Drug Reaction Monitoring, Chongqing, China

[3]Chongqing Academy of Science & Technology, Chongqing, China

Corresponding Author:
Daidi Zhong, PhD
Bioengineering College
Chongqing University
Shazheng Street 174
Shapingba
Chongqing, 400044
China
Phone: 86 2365102507
Email: daidi.zhong@hotmail.com

Abstract

Background: The adverse event report of medical devices is one of the postmarket surveillance tools used by regulators to monitor device performance, detect potential device-related safety issues, and contribute to benefit-risk assessments of these products. However, with the development of the related technologies and market, the number of adverse events has also been on the rise, which in turn results in the need to develop efficient tools that help to analyze adverse events monitoring data and to identify risk signals.

Objective: This study aimed to establish a hazard classification framework of medical devices and to apply it over practical adverse event data on infusion pumps. Subsequently, it aimed to analyze the risks of infusion pumps and to provide a reference for the risk management of this type of device.

Methods: The authors define a general hierarchical classification of medical device hazards. This classification is combined with the Trace Intersecting Theory to form a human-machine-environment interaction model. Such a model was applied to the dataset of 2001 to 2017 class I infusion pump recalls extracted from the Food and Drug Administration (FDA) website. This dataset does not include cases involving illegal factors.

Results: The proposed model was used for conducting hazard analysis on 70 cases of class I infusion pump recalls by the FDA. According to the analytical results, an important source of product technical risk was that the *infusion pumps did not infuse accurate dosage (ie, over- or underdelivery of fluid)*. In addition, energy hazard and product component failure were identified as the major hazard form associated with infusion pump use and as the main direct cause for adverse events in the studied cases, respectively.

Conclusions: The proposed human-machine-environment interaction model, when applied to adverse event data, can help to identify the hazard forms and direct causes of adverse events associated with medical device use.

(JMIR Hum Factors 2019;6(2):e10366) doi:10.2196/10366

KEYWORDS

infusion pump; risk management; equipment failure; hazard analysis and critical control points; man-machine systems

Introduction

Infusion Pumps

Continuous intravenous delivery of drugs with short half-lives, such as inotropic agents and vasodilators, is a recommended technique in acute care [1]. A syringe pump is a device that intravenously infuses fluids, drugs, or nutrients in the patient [2]. The use of infusion pump is helpful as it helps in reducing nurses' workload and in improving accuracy and efficiency in terms of delivery of drugs or fluids. The purpose of using a syringe pump in clinical settings is to administer an accurate

amount of drug or fluid over a relatively long duration, and it can be especially favorable for continuous infusion of very small amounts such as 0.1 mL/hour [3]. In clinical settings, transfusion pumps and syringe pumps are referred to as *infusion pumps*.

The infusion pump system is mainly composed of the following parts: the microcomputer system, the pump component, the detection system, the alarm device, and the input and display devices. The microcomputer system controls and manages the whole system intelligently, prevents the occurrence of incorrect infusion, and sends an alarm signal to the alarm device for sound and light alarms. The pump component is the power source of the liquid injection. The detection system, which is usually made up of different kinds of sensors in different parts, is used to detect the working state of the infusion pump, thereby facilitating the detection of all kinds of abnormalities in time. The alarm device is used to inform the medical and nursing staff of the normal and abnormal states. The input part is used to set the parameters of the infusion, such as the amount of infusion and the speed of the infusion. The display section is responsible for displaying the parameters and the current working state.

The use of infusion pumps was identified as the area with the highest risk, based on incident report data [4]. In unique studies by Keers et al, a higher median medication administration error (MAE, 10 studies used denominators falling within the total opportunity for error [TOE] definition of the 12 studies that examined only intravenous administration) rate was observed for the intravenous route (53% excluding timing errors; interquartile range [IQR]: 27%-58%) without timing errors (n=5) using tolerable negative error (TNE) compared with when all administration routes (56 used the TOE denominator of the 61 studies observing all routes of administration) were studied (20%; IQR: 9%-25%) without timing errors (n=17) using TNE, where each dose could accumulate more than 1 error [5]. Intravenous infusion may present the greatest preventable MAE risk to hospitalized patients [6]. At present, infusion pumps are commonly used in clinics; however, problems exist with respect to the use of infusion pumps in clinics, including discontinued infusion, leakage, inaccuracy of infusion dose, and too fast or too slow infusion speed. According to the clinical needs, analyzing the failure and mode effect of infusion pumps was useful for evaluating the ease of use and ergonomics and evidence-based procurement [7].

The failure modes and infusion errors of infusion pumps are always the top 10 hazards on ECRI Institute's annual list. Top-ranked hazards of 2017 announced by ECRI Institute focus on infusion errors that can occur when using large-volume infusion pumps [8]. On August 23, 2013, ECRI Institute Patient Safety Organization's (PSO) clinical engineering staff found certain risks associated with use of infusion pumps during a regular review of device-related events submitted to the PSO. The team saw multiple events at 1 hospital in which an infusion pump had stopped working with no apparent cause. Investigation revealed that a disconnection between the pump module and the personal computer (PC) unit had caused unexpected cessation of infusion therapy for several patients. The problem resulted from corroded or damaged interunit interface connectors [9]. In the ECRI Institute's PSO Monthly Brief published in February 2015, the patient safety analyst of ECRI Institute PSO, Stephanie Uses, emphasized the potential risk on each phase of the medication use process. She said that there is a risk of confusion among look-alike or sound-alike injection drugs formations, concentrations, and dosages when prescribing the proper one for the patient during the prescribing stage. Risks during the monitoring phase include inadequate monitoring—when patients' response to insulin is not observed to see if an adjustment in dose is necessary [10].

Thus, effectively decreasing the risks of infusion pumps in clinical settings will be critical for improving the success rate for emergency treatment of patients. In 2010, Zhang et al introduced a generic insulin pump model and a preliminary hazard analysis based on this model [11]; they divided the hazardous situations into 5 categories associated with the generic insulin infusion pump, including therapeutic, energetic, chemical or biological, mechanical, and environmental. Curzon et al established a tool focused on understanding how the design of interactive medical devices (such as infusion pumps, monitors, and diagnostic devices help save lives) can support safety [12]. Masci presented a hazard analysis that identified a substantial set of root causes of use hazards in software design, which is general in the sense that the problematic functionalities are common in broad classes of infusion pumps [13]. Masci et al established a model-based risk analysis methodology that helps manufacturers identify and mitigate use hazards in their products at early stages of the development life cycle [14]. They also presented a generic user interface (UI) architecture, Generic Infusion Pump User Interface, to facilitate the identification and reasoning of use hazards in infusion pumps [14].

Medical Device Adverse Events

The medical devices, because of their natural characteristics, may bring safety risks, together with health benefits, to the users. The adverse events associated with the qualified postmarket medical devices cause a variety of harms (or potential harms) to the human body under normal operation. These adverse events (including any symptoms, signs, diseases, or the events could result in significant injury or death) do not necessarily have a direct causal relationship with medical devices and can only be temporarily associated with medical devices. The monitoring of medical device adverse events can be useful in warning health care institutions and regulatory bodies on how to use medical devices safely and effectively. All national regulators have established a corresponding data reporting system to actively collect medical device adverse events.

To reduce or avoid the possible risks and damage to human health caused by medical devices, recalling the postmarket defective medical devices is an internationally accepted method for safety management of postmarket medical devices. Being one of the active practitioners of medical device recall, the US Food and Drug Administration (FDA) categorizes all recalls into 3 classes according to the level of hazards caused by medical devices. The class I is defined as dangerous or defective products that predictably could cause serious health problems or death [15]. The recalls are available in the Medical

Devices/Safety/List of Recalls on the FDA's official website [16].

In 1972, Professor Elwyn Edwards first proposed the principle of *human* as the center of a particular system interface in security work, including elements such as software, hardware, environment, and liveware (SHEL). The acronym SHEL stands for these 4 elements, and these factors constitute the SHEL model. The human error should be analyzed because of the mismatch between interfacial elements. With respect to the use of medical device risk analysis, in 2011, Long et al established a *medical personnel* –centric medical device risk analysis model based on the SHEL model, called device, environment, liveware, patient, software [17]. Masci et al presented a hazard analysis method that extends Leveson's System Theoretic Process Analysis with a comprehensive set of causal factor categories so as to provide developers with clear guidelines for systematic identification of use-related hazards associated with medical devices, their causes embedded in UI software design, and safety requirements for mitigating such hazards [18]. Harrison et al concerned with how to demonstrate that a UI software design is compliant with use-related safety requirements, and they established a methodology that aims to demonstrate how to achieve the FDA's agenda of using formal methods to support the approval process for medical devices [19]. Masci et al established a technique that integrates human cognitive process models and general interaction design principles and uses a model-based approach for systematic exploration of potential hazards [20].

However, from the perspective of medical device supervision, the goal of postmarket medical device risk management is to further discover the causes of unacceptable risks associated with medical devices products through production and postproduction safety data (including medical device adverse events), such as product design, production process, specifications and other issues, and then take appropriate risk control measures, that is, considering *product* as the center of the risk analysis, carrying out evaluation and control process, and making sure that its starting point and foothold are *products*.

Therefore, based on the above research results, this paper presents a hazard classification framework of the medical devices and human-machine-environment interaction model, which was used to analyze 70 cases of FDA class I infusion pump recalls, to identify the direct cause of all risks, then put forward some advice for the life cycle management of infusion pumps.

Methods

Overview

Adverse event reports are the main source of data for this study. Our aim here was to find key hazard risk factors and direct causes through the analysis of adverse event reports. Analyzing the hidden risk of medical device based on adverse event report is generally considered as a complicated job. The risk factors cannot be directly extracted if we do not have an appropriate tool to structuralize the content in those reports. For example, in an infusion pump, the application environment is a complex system of human-machine-environment interaction. It is almost impossible for us to identify the hidden risk factors without thoroughly understanding this complex system. Therefore, in this study, we developed a tool that allows the modeling of such a complex system, and then, we used this tool to analyze the hazard of infusion pumps.

This tool was developed based on the Trace Intersecting Theory, which is a widely used generic tool for the analysis of a complex system. However, it is too general to be directly applied to our target—infusion pumps. To better adapt to the characteristics of medical device products, we extended this theory with 5 new types, so that the model could be applied to the risk analysis of medical devices, and then applied it to analyze the infusion pump recalls.

The Hazard Classification Framework

In terms of the evolution process of safety theory, the early theories of *accident proneness* emphasized the influence of people's personality characteristics on accidents. Later in 1931, Heinrich put forward the *accident causation theory*, emphasizing that accidents are the result of the interaction of various factors. In 1961 and 1966, Gibson and Haddon introduced a new concept: accidents are incorrect or undesirable energy transfers or releases. At this time, it was discovered that injury accidents could be prevented by controlling energy. In 1969, Surry suggested that people's mishandling of information might lead to accidents. After the inheritance and development of these ideas by many people, it was found that the unsafe behavior of people or the unsafe state of things is the direct cause of industrial accidents.

The Trace Intersecting Theory focuses on the cause of the accident. Such causes can be summarized as equipment's faults (or defects) and human errors. The intersection of the 2 event chains indicates an accident. The basic idea is that injury accidents are the result of the development of 2 series of interrelated people and things (including the environment). As a result of a variety of factors, the unsafe behaviors of people and the unsafe state of the objects will keep on evolving in their respective trajectories, and accidents will happen at a later point when they meet or interact at a certain time and space (see Figure 1).

The occurrence mechanism of medical device adverse events consists of 4 types of interactive factors (see Figure 2). Among them, the *parasitifer* is an individual who may be injured, including the patient and/or medical personnel. The *applicator* is the medical device that generates force and transmits or prevents energy, and a *human-machine* relationship is formed between the *applicator* and the *parasitifer*. For the purpose of diagnosis and treatment, the exchange or transmission of material, energy, and information between the human body system and the medical device system will continue. When the material, energy, and information involved in the exchange or transmission exceed the limit tolerance of the human body, a certain type of harm will occur, which we refer to as *hazard mediums*. The *hazard situation* focuses on the conditions or environment in which the injury occurs, that is, the condition and degree of the human body in various hazardous environments.

Figure 1. The schematic diagram of the Trace Intersecting Theory.

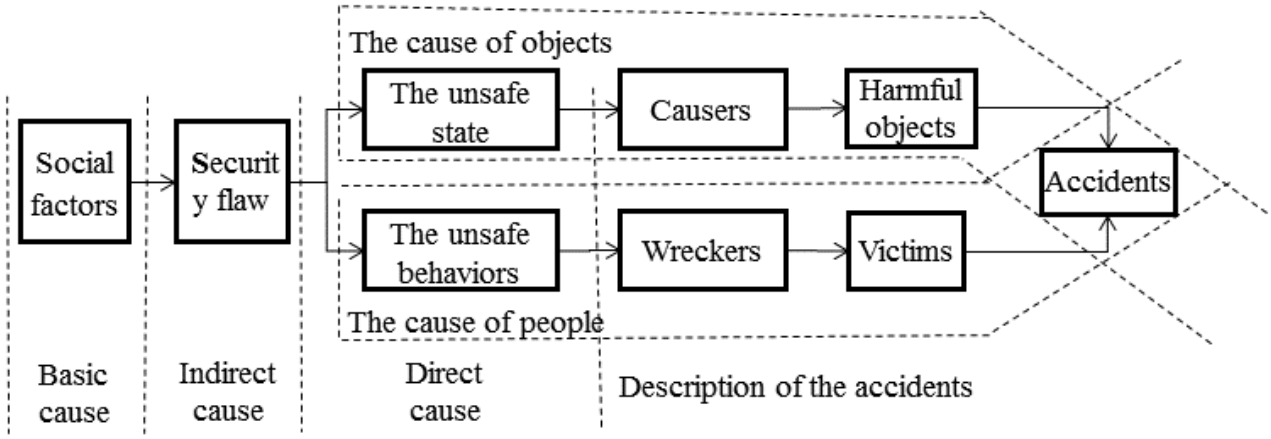
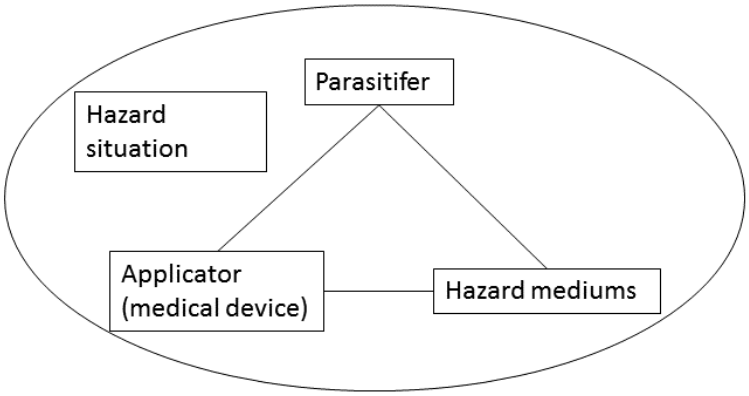


Figure 2. The mechanism of medical device adverse events.



Modern physics considers that material and energy are the elements of the objective world, but a closer look will find that information is another attribute of the objective world, in addition to the material and energy [21]. Therefore, we use the material, energy, and information as the 3 fundamental elements to model the objective world for the purpose of hazard analysis, but it is difficult to separate material from energy because energy exists in any type of material, and energy cannot live alone without the material being its host. Thus, in the following analysis, the material and energy are merged together and is analyzed as *energy*.

Therefore, medical device adverse events can be divided into 3 types based on different hazard mediums: (1) Type I: Energy hazard; (2) Type II: Information hazard; and (3) Type III: Energy and Information hazard (see Table 1).

The Energy hazard medium is called type I medical device adverse event. It refers to the event wherein medical devices may directly cause human injury in the form of energy under the application environment [22]. The Energy hazard can be further divided into 2 subtypes: the excess energy and the insufficient energy. Among them, the excess energy refers to the scenario when certain kind of energy exceeds the threshold that the humans can bear, which may directly or indirectly lead to the damage of human body. The form of such excess energy can be mechanical energy (Ia-01), radiant energy (Ia-02), thermal energy (Ia-03), electricity (Ia-04), biological and chemical energy (Ia-05), and the others (Ia-06). The insufficient energy refers to an event that may cause human injury directly because of interference in the normal life energy and material exchange, between the human body and the surrounding environment. These cases are in the form of hypoxia, hypothermia, and hydropenia, which can cause exchange impairment between the human body and the surrounding environment (Ib-01), or the failure of life support or first-aid in critically ill patients (Ib-02), and the others (Ib-03).

The Information hazard is called a type II medical device adverse event. It refers to events that may directly cause human injury in the form of information under the application environment. This type of hazard can be further divided into 3 types: incorrect information, insufficient information, and overloaded information, which are in the form of data, text, sound, and image.

The Energy and Information hazard has the characteristics of both the type I and type II hazards at the same time and is called type III medical device adverse event. According to the weight of each constitutional hazard, the type III hazard can be divided into 3 subtypes: the energy-dominant, information-dominant, and dual-culprit. The dual-culprit subtype means that both Energy and Information contribute significantly to the hazard.

Table 1. The hazard classification framework of the medical devices.

Hazard classification	Subtype
Type I: Energy hazard	Subtype Ia: (excess energy)
	Subtype Ib: (insufficient energy)
Type II: Information hazard	Subtype IIa: (incorrect information)
	Subtype IIb: (insufficient information)
	Subtype IIc: (overloaded information)
Type III: Energy and Information hazard	Subtype IIIa: (energy-dominant)
	Subtype IIIb: (information-dominant)
	Subtype IIIc: (dual-culprit)

The Direct Causes Classification

From the viewpoint of system security, the risk factors of *human-machine-environment* system come from 3 interrelated aspects: *human*, *machine*, and *environment*. In a specific environment, the user has acquired recognition, perception of different information of medical devices, and repeated the actual operation. Through this process, medical devices can be controlled and used to diagnose and treat patients. To describe how a hazard was caused by such interaction among human, medical device, and environment, the authors define a human-machine-environment interaction model (see Figure 3) that contains 5 kinds of direct causes (operator-device, O-D; patient-device, P-D; environment-device, E-D; device, D; and unknown, U). Each direct cause (see Table 2) represents a set of direct causes of certain group of adverse events.

Figure 3. The schematic diagram of the human-machine-environment interaction model.

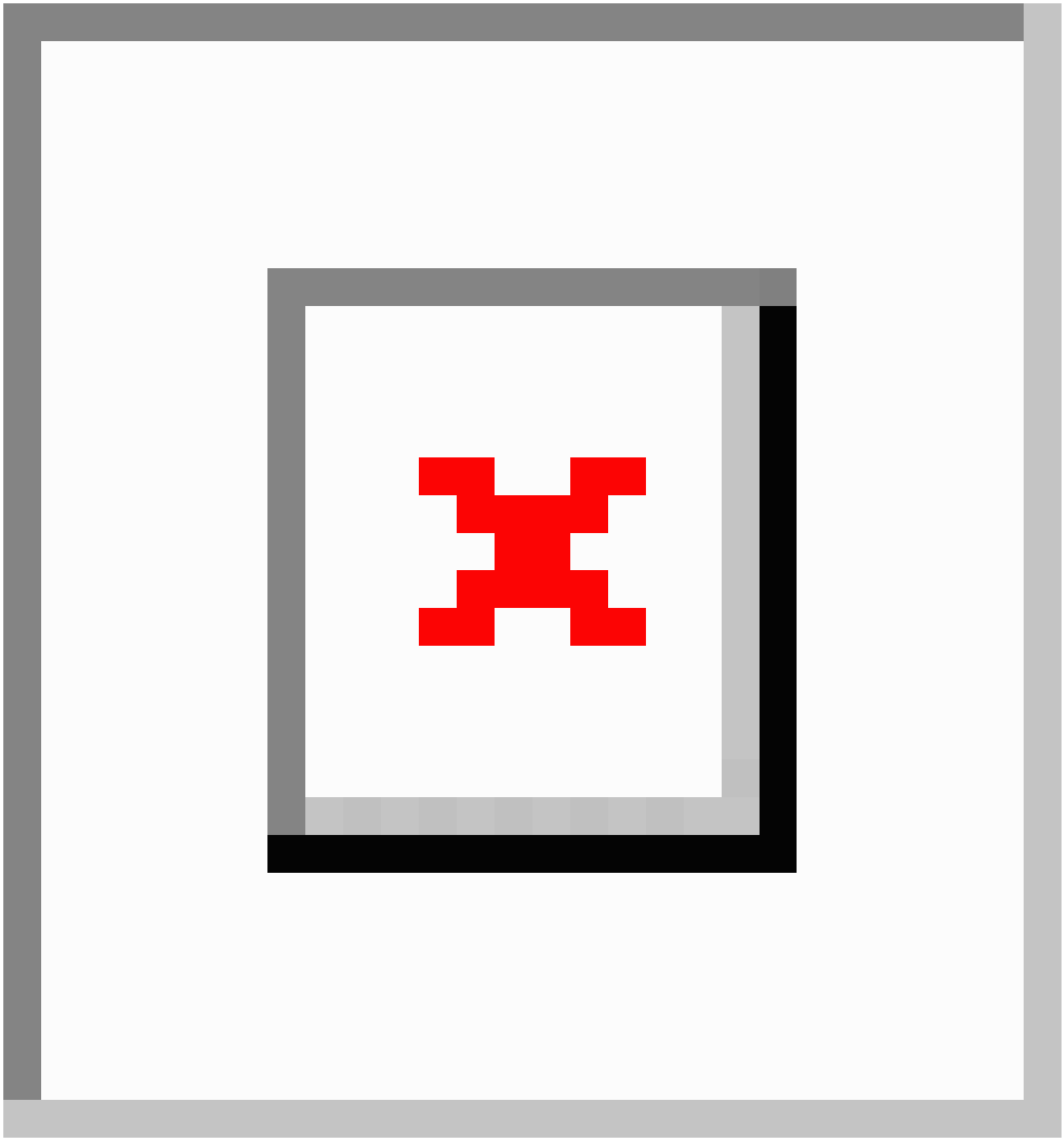


Table 2. The type of direct causes.

Direct cause	Description	Main forms
Operator-device	A safety accident that may be caused by a problem in the interaction between the operator and the device	Usability problems: display interface; control interface; HMI[a] matching (space, seat, and workspace); label or specification; other
Patient-device (P-D)	A safety accident that may be caused by a problem in the interaction between the patient and the device	P-D-1 usability problems: display interface; control interface; HMI matching (space, seat, and workspace); label or specification; other
		P-D-2 internal risk: biocompatibility (blood, tissue, and immunoreaction); tissue or organ infection; tissue or organ damage; other
Environment-device	A safety accident that may be caused by a problem in the interaction between the environment and the device	Effect on the environment or disturbance by other devices: pollution (eg, air pollution); disturbance (eg, electromagnetic interference; other
Device (D)	A safety accident that may be caused by component failure of the device	D-1 (hardware failure); D-2 (software failure)
Unknown	Unknown causes or unexpected injuries	Unknown scientific principle involves multiple chaotic factors and unexpected events

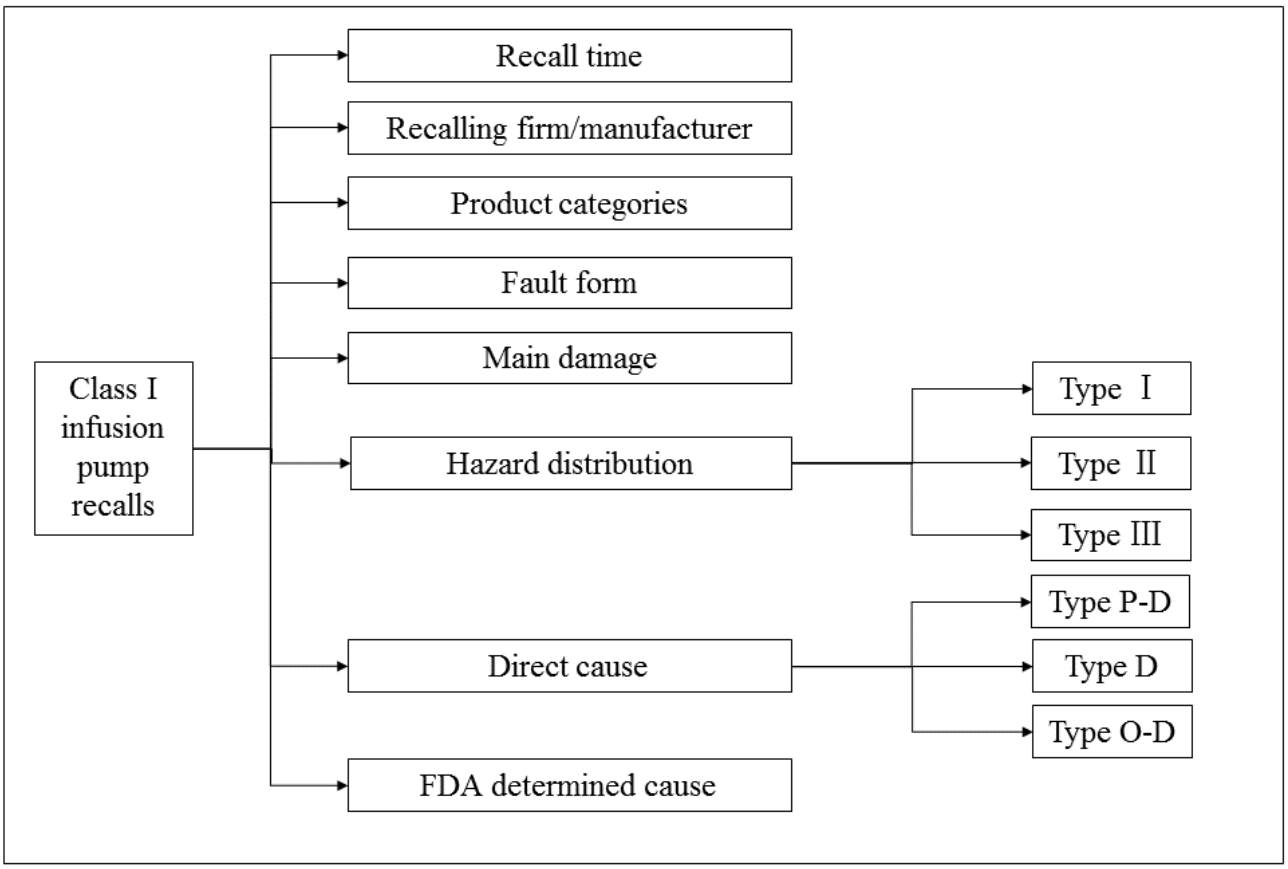
[a]HMI: human-machine interface.

The O-D type direct cause refers to the safety events caused by the problem in interaction between the operator and the device, which is mainly expressed as the availability problems, including display interface, control interface, and label or specification. The P-D type direct cause refers to the safety events caused by the problem in the interaction between the patient and the device, which is mainly expressed as the availability problems and the internal risk. The interpretation of the availability problems is the same as above. The internal risks include biocompatibility (blood, tissue, and immune response), tissue or organ infection, and tissue or organ injury. The E-D type direct cause refers to the safety events caused by the interaction between the environment and the device, which is mainly expressed as the equipment affecting the work environment or being affected by other facilities. The D type direct cause refers to the safety events caused by the failure of the device component, which is mainly expressed as hardware failure and/or software failure. The U type direct cause refers to the safety events caused by unknown causes or unexpected injuries. Among them, O refers to operator, P refers to patient, D refers to device, E refers to environment, and U refers to unknown.

To help readers to better understand the use of the hazard classification framework established in this paper, the following example provides detailed instructions, shown in Multimedia Appendix 1: ID 17. Manufacturer reason for recall: package labeled as an insulin syringe for use with U-100 insulin contains an insulin syringe for use with U-40 insulin. This entails the risk of overdose of insulin. The incident involved 2 aspects of the hazard, including overdose of insulin (Ia-05) and mislabeling (IIa), which is caused by the problem in interaction between the operator and the device (the O-D type direct cause).

Figure 3 illustrates the pathway of performing statistical analysis over infusion pump recall by leveraging the above human-machine-environment interaction model and the hazard classification framework (see Figure 4).

Figure 4. The inferencing pathway of statistical analysis. FDA: Food and Drug Administration. O-D: Operator and device. P-D: Patient and device. D: Failure of the device.



Results

The Basic Information Statistics

Figure 5 shows the number of class I infusion pump recalls released by FDA from 2001 to 2017. The largest number of recalls occurred in 2013, which accounted for 20% of the total. The number of recalls from 2001 to 2006 showed a rising trend, and thereafter, a downward trend was observed for 5 years after 2006. The total number of recalls from 2012 to 2015 accounted for 53% of the total, and there was a gradual decline trend after 2013.

Product recalls were mainly issued by the following firms or manufacturers: Medtronic Inc, Hospira Inc, Baxter Healthcare Corp, and CareFusion 303, Inc (see Table 3). The total number of recalls for the 4 companies accounted for 57%. However, the largest number of recalls of a company's products does not indicate that the company's products are more risky, because a bigger market share is likely to increase the number of recalls.

Infusion pumps can be divided into the following subtypes: injection pump, elastic pump, and peristaltic pump [23,24]. The most common type of injection pump is the insulin pump. The nutrition pump is an example of the peristaltic pump, and the disposable infusion pump is an example of an elastic pump [23]. Infusion pumps are also categorized into epidural pumps and intravenous pumps. The epidural pump is a topical medication, and the intravenous pump is a systemic medication; the epidural pump can achieve a good analgesic effect with very few drugs, but the catheter is easy to fall off when the patient moves.

Of the 70 cases, 12 (17%) are passive devices, including 6 cases of disposable medical equipment and 6 cases of infusion pump components. The remaining 58 (83%) are active equipment. Infusion pumps make up the maximum proportion, followed by insulin infusion pumps. There were 7 cases of recalls related to infusion pump applications (see Table 4).

There were 17 cases of adverse events caused by *software failures* (see Table 5). The main outcome of equipment faults was product component failure, characterized by sensor failure, pump door breakdown, flow restrictor failure, keypad failure, infusion tube bending or occlusion, the detachment of catheter access port from the main body of the pump, etc. As shown in Table 5, there were many occurrences of power failures and alarm failures (no alarm and false alarm). Furthermore, there can be other problems such as mislabeling, backflow or free flow, and unintended higher flow rate.

Of the 70 cases, 66 described the main damage to patients (see Table 6), which manifested as infection, overdose, underdose, and incorrect treatment. It is known that underdose can result in delay or interruption of infusion therapy, serious injury, or death. Moreover, a drug overdose can lead to serious adverse clinical consequences such as respiratory depression, coma, or death.

Figure 5. The distribution of recall time.

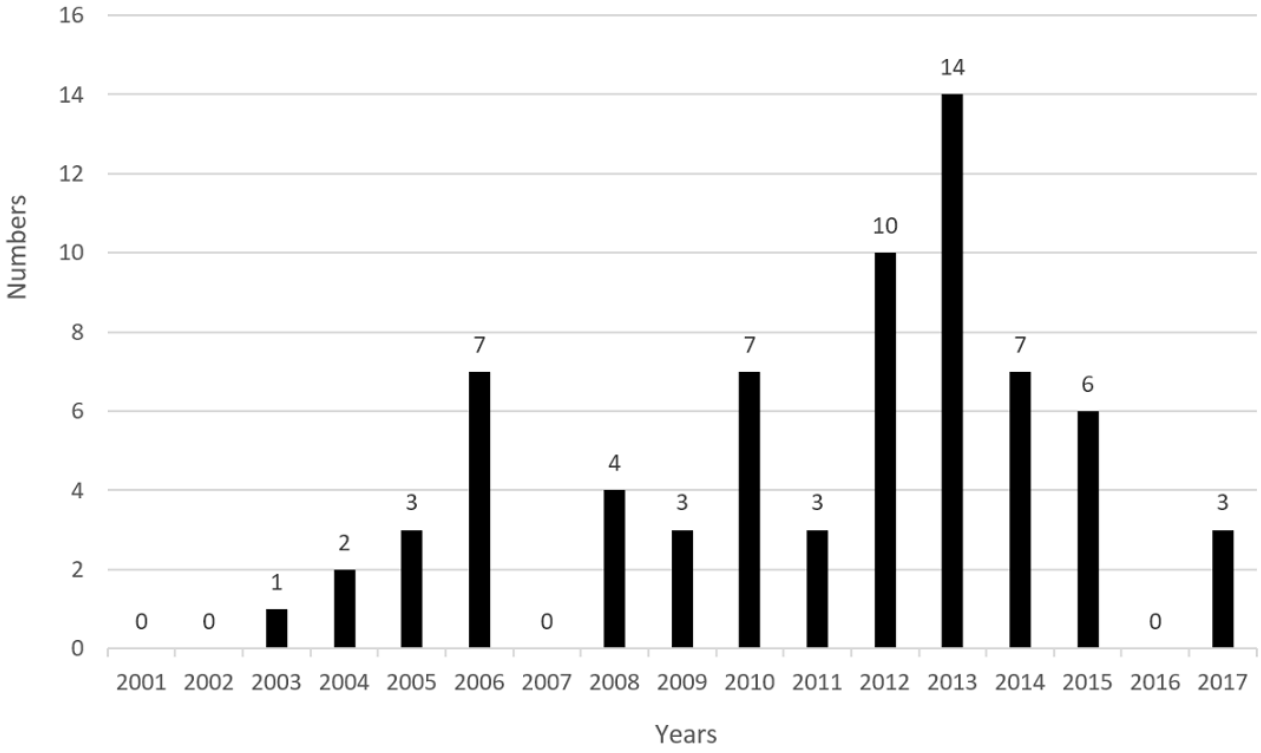


Table 3. The distribution of recalling firm or manufacturer.

Recalling firm/manufacturer	Recalls, n
Medtronic Inc	14
Hospira Inc	11
Baxter Healthcare Corp	8
CareFusion 303, Inc	7
Disetronic Medical Systems, Inc	3
Animas Corporation	2
B. Braun Medical, Inc	2
Cardinal Health	2
Covidien	2
Insulet Corporation	2
Sigma International General Medical Apparatus, LLC	2
Codman & Shurtleff, Inc.	1
Elite Biomedical Solutions LLC	1
First Medical Source LLC	1
ICU Medical, Inc	1
I-Flow Corporation	1
Iradimed Corporation	1
Manufacturer Codman & Shurtleff, Inc	1
Micromedics, Inc	1
MOOG Medical Devices Group	1
Nurse Assist, Inc	1
Roche Insulin Delivery Systems Inc	1
Smiths Medical ASD, Inc	1
Symbios Medical Products, LLC	1
Tandem Diabetes Care Inc	1
Walkmed Infusion LLC	1

Table 4. The list of product categories (N=70).

Product categories	n (%)
Intravenous injection transfusion system	4 (6)
Infusion pump applications	7 (10)
Insulin infusion pump	12 (17)
Infusion pump	47 (67)

Table 5. The list of the fault form. A case of a recall may have multiple equipment failures.

Equipment faults	Count, n
Electrical shorting	1
Failure to detect air-in-line conditions	1
Weak seals of the sterile pouches	1
Cartridges leaking	1
Mislabeled	2
Unexpected shutdown	2
Higher flow rate	3
Backflow or free flow	3
Power failure	9
Alarm failures	15
Software failures	17
Component failure	22

Table 6. The list of the main damage (N=66).

Hazard	n (%)
Infection	3 (5)
Incorrect treatment	9 (14)
Overdose	18 (27)
Underdose	36 (55)

The Hazard Classification Statistics

Next, we examined the effect of the hazard classification framework. These data suggest that Energy hazard was the major form of expression (see Table 7).

Certain cases of subtype I hazard may correspond to multiple harmful mediums form, which could be recognized as both the excess energy and insufficient energy. Due to this, the 47 cases of type I hazard in Table 7 actually contain 27 cases of excess energy and 32 cases of insufficient energy (Table 8 shows the corresponding detailed distributions).

The results show that the subtype II hazard (Information hazard) includes 1 case of incorrect information and 3 cases of insufficient information. Moreover, 19 cases of subtype III hazard (Energy and Information hazard) include 14 cases of energy-dominant and 5 cases of information-dominant.

Table 7. The distribution list of hazard distribution (N=70).

Hazard classification	n (%)
I	47 (67)
II	4 (6)
III	19 (27)

Table 8. The distribution list of type I. A case of a recall may have multiple hazard classifications.

Subtype and energy medium	n
Ia (n=27)	
I a-01	1
I a-03	1
I a-05	25
Ib (n=32)	
I b-02	25
I b-03	7

The Direct Causes Classification Statistics

Finally, we carried out a statistical analysis on the direct cause. There were 72 cases by reason of a case of a recall that may have multiple direct causes. As shown in Table 9, the D type direct cause makes the maximum proportion.

The availability issues can be observed from the O-D type direct cause, including 2 cases of mislabeled and 4 cases of control interface problems. The D type direct cause includes 17 cases of software failures and 45 cases of hardware failures (see Table 10).

We have noticed that the FDA website published the FDA-determined cause. The statistical analysis revealed device design to be the main cause (see Table 11).

Table 9. The distribution list of direct cause (N=72).

Direct cause	Statistics, n (%)
Patient-device	2 (3)
Operator-device	6 (8)
Device	64 (89)

Table 10. The distribution list of the D type direct cause (N=65).

The D type direct cause	n (%)	Event manifestations
Software failures	17 (26)	Unexpected shutdown, communications errors
Hardware failures	45 (69)	Component failure, material fracture
Invalid information	3 (5)	__[a]

[a]It was difficult to judge if the main form of the case is hardware failure or software failure, but it was certain that it was caused by the component failure of the device.

Table 11. The distribution list of Food and Drug Administration–determined cause.

Food and Drug Administration–determined cause	n
Equipment maintenance	1
Labeling design	1
Mix-up of materials or components	1
Packaging process control	1
Pending	1
Process change control	1
Software manufacturing or software deployment	1
Use error	1
Component change control	2
Under investigation by firm	2
Component design or selection	3
Process control	3
Process design	4
Nonconforming material or component	6
Software design	6
Other	7
Device design	28

Discussion

Overall, our study established a hazard classification framework for medical devices. Through the statistical analysis on the above 70 cases of FDA class I infusion pump recalls, our results confirmed that the key contributor to the product technical risk is that the *infusion pump did not infuse accurate dosage (over- or underdelivery of fluid).*

Product Component Failures

Most product component failures are caused by device design. The most popular cases within this type of failure are listed below:

1. The *sensor failure* may generate a false alarm or an undetected fluid buildup within the distal line, resulting in delay or interruption of therapy or overinfusion of fluid.
2. The *full or partial occlusion of the infusion tube* may prevent fluid from reaching the patient, causing an interruption of delivery.
3. The normal closure of the *pump door* is closely related to the dosage delivered, which helps the patient to ensure proper treatment process. If the door assembly breaks, it may prevent the door from closing properly; thus, unrestricted flow may occur. If the door cannot be closed, the pump cannot be used, and this will lead to a delay in therapy.
4. The displacement of the *Flow restrictor bead* may be the root cause of the fast flow of contents.
5. The *Luer tube may break at the connection to the pump*, and if this is not noticed by the patient, the patient may receive an underdelivery of drug. A delay or interruption in therapy has the potential to result in a worst-case scenario such as significant injury or death. Similarly, depending on the drug and the dosage delivered, overinfusion also has the potential to result in significant injury or death.
6. In addition, one fact that may explain these defects is that some companies start selling these pumps when they are still in research and development. This was typically the case for Hospira with the Symbiq pump.

Software Failure

Of the 70 cases of adverse events (Table 10), 17 were caused by *software failures*. Such failures are usually characterized by the following adverse event contents: *wrong instruction*, *error codes*, or *communication errors*. The operator may execute the wrong operation according to the wrong instruction, resulting in an overdose or underdose.

Alarm Failures

Of the 70 cases of adverse events, 15 were caused by alarm failures, including 5 cases of *false alarm* and 10 cases of *no alarm*. The main forms include (1) pump shutting off during use without warning and (2) *a false visual or audible alarms* causing the infusion pump to stop supplying the fluids to the patient. The fault alarm system may be due to the failure of *internal detector, inability to trigger the alarm*, *the fault of software*, or *lack of regular maintenance*. Alarm hazards are among the top 5 hazards on ECRI Institute's 2011 list [25]. These studies could help hospitals to enhance their management system, for example, to improve the existing nurse training system, thus to better educate nurses about their shared responsibilities. At the same time, these studies also provide a new strategy to ensure the safe use of medical devices. Nurses should not only pay attention to the operation procedures but also focus on maintenance. In fact, the shortage of nurses is another possible reason for the failure to maintain medical devices. More importantly, manufacturers can also strengthen postmarket maintenance.

Power Failure

Power failure can result in the situation where the device ceases operation without warning and also loses the data. An incorrect voltage could potentially lead to a loss of communication between the PC unit main processor and the keyboard processor, which can lead to unexpected loss of therapy. Excessive battery discharge can damage the batteries and may further interrupt the therapy. Therefore, we recommend manufacturers to consider designing other backup power and to simplify the operation of replacing batteries.

Altogether, product component failure is the main direct cause of the infusion pump failure. The Energy hazard, containing the excess energy subtype and insufficient energy subtype, is the major form of the hazard of the infusion pump. Among the excess energy type of hazards, *infection* and *overdose* occur most frequently, but the *interruption of infusion therapy* is the hazard that causes the most serious injuries. A substantial part of the hazard of insufficient energy is the *interruption of therapy*, which is mainly caused by *unexpected shutdown*, *power failure*, or *component failures*.

Limitations of This Research

The biggest problem is that manufacturers, distributors, medical institutions, and device users fail to actively cooperate with the supervision department. Moreover, many steps should be performed by health care institutions before implementing a pump, which can avoid some of the problems faced by infusion pump users. In particular, many defects are not reported to the FDA or other agencies (eg, Health Canada) but directly to the providers of infusion pumps. As a result, many other types of events are not reported, for example, free flow, valve dysfunction, foam in the product because of the mechanism of the pump, and hemolysis. Therefore, there is a lack of sufficient data to further optimize the model in the research work. In addition, influential factors such as the service life of medical devices does not appear in the report, which in turn increases the difficulty of the research.

Conclusions

With social progress and development of technology, infusion pumps are widely being used in clinical settings. There is a potential safety risk while alleviating the patient' s suffering, so it is of great significance to ensure proper and safe use of infusion pumps. This study investigated the direct cause of occurrence of infusion pump risks. This may help to provide reference for the infusion pump risk management as well as effective information for safe use and infusion pump safety in clinical environments. To this end, we propose a new data analysis method that can help reveal a single type of adverse events' risk characteristics and common problems of medical devices based on the Trace Intersecting Theory. It can be used to guide the specific quality monitoring work for the FDA and national authorities to form a complete regulatory system for postmarket medical devices.

We believe that carrying out risk assessment and analysis work for the postmarket medical devices is of great significance, which can optimize the product risk control solutions and have a positive effect on the development of public health.

Acknowledgments

This paper was partially funded by Chongqing University via the Ministry of Science and Technology project (#2018YFC2000804). As the sponsor of this study, Chongqing University was not involved in the review and approval of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The hazard classification of 2001 to 2017 class I infusion pump recalls extracted from the Food and Drug Administration website.

[PDF File (Adobe PDF File), 316KB - humanfactors_v6i2e10366_app1.pdf]

References

1. Weiss M, Bänziger O, Neff T, Fanconi S. Influence of infusion line compliance on drug delivery rate during acute line loop formation. Intensive Care Med 2000 Jun;26(6):776-779. [doi: 10.1007/s001340051246] [Medline: 10945397]
2. Jung B, Seo KS, Kwon SJ, Lee K, Hong S, Seo H, et al. Efficacy evaluation of syringe pump developed for continuous drug infusion. J Dent Anesth Pain Med 2016 Dec;16(4):303-307 [FREE Full text] [doi: 10.17245/jdapm.2016.16.4.303] [Medline: 28879319]
3. Jacobs B. Using an infusion pump safely. Nursing 2006 Oct;36(10):24. [Medline: 17019326]
4. Fechter RJ, Barba JJ. Failure mode effect analysis applied to the use of infusion pumps. Conf Proc IEEE Eng Med Biol Soc 2004;5:3496-3499. [doi: 10.1109/IEMBS.2004.1403981] [Medline: 17271040]
5. Keers RN, Williams SD, Cooke J, Ashcroft DM. Prevalence and nature of medication administration errors in health care settings: a systematic review of direct observational evidence. Ann Pharmacother 2013 Feb;47(2):237-256. [doi: 10.1345/aph.1R147] [Medline: 23386063]
6. Elias BL, Moss JA. Smart pump technology: what we have learned. Comput Inform Nurs 2011 Apr;29(4 Suppl):TC61-TC67. [doi: 10.1097/NCN.0b013e31821ef813] [Medline: 21562383]
7. Poder TG. Using the health technology assessment toolbox to facilitate procurement: the case of smart pumps in a Canadian hospital. Int J Technol Assess Health Care 2017 Jan;33(1):54-62. [doi: 10.1017/S0266462317000125] [Medline: 28578750]
8. ECRI Institute. 2017 Jan. 2017 Health Technology Hazards Topped by Infusion Errors URL: https://www.ecri.org/EmailResources/PSO_Monthly_Brief/2017/PSO_Brief_0117.pdf [accessed 2019-03-12] [WebCite Cache ID 76otedcSe]
9. ECRI Institute. PSO Prevents Safety Risk with Infusion Pumps URL: https://www.ecri.org/Pages/PSO-Prevents-Safety-Risk-with-Infusion-Pumps.aspx [accessed 2018-03-10] [WebCite Cache ID 6xoj1TUvu]
10. ECRI Institute. Insulin Errors Occur Across the Medication Use Process URL: https://www.ecri.org/EmailResources/PSO_Monthly_Brief/2015/PSO_Brief_0215.pdf [accessed 2019-03-12] [WebCite Cache ID 76otptNAu]
11. Zhang Y, Jones PL, Jetley R. A hazard analysis for a generic insulin infusion pump. J Diabetes Sci Technol 2010 Mar 1;4(2):263-283 [FREE Full text] [doi: 10.1177/193229681000400207] [Medline: 20307387]
12. Paul C, Ann B, Harold T, Anna C. Safer interactive medical device design: Insights from the CHI+MED Project. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare. 2015 Presented at: MOBIHEALTH'15; October 14-16, 2015; London, Great Britain. [doi: 10.4108/eai.14-10-2015.2261752]
13. Masci P. HASLab: INESC TEC. Technical report: A preliminary hazard analysis for the GIP number entry software URL: https://haslab.uminho.pt/masci/files/techrep-pha.pdf [accessed 2019-03-15] [WebCite Cache ID 76srZBIsZ]
14. Masci P, Zhang Y, Jones P, Thimbleby H, Curzon P. A Generic User Interface Architecture for Analyzing Use Hazards in Infusion Pump Software. In: Proceedings of the 5th Workshop on Medical Cyber-Physical Systems. 2014 Presented at: 5th Workshop on Medical Cyber-Physical Systems; April 14, 2014; Berlin, Germany p. 1-14. [doi: 10.4230/OASIcs.MCPS.2014.1]
15. US Food and Drug Administration. FDA Product Recalls-From First Alert to Effectiveness Checks EB/OL URL: https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm049070.htm [accessed 2018-03-10] [WebCite Cache ID 6xoj8FjhU]
16. US Food and Drug Administration. URL: https://www.fda.gov/ [accessed 2019-03-18] [WebCite Cache ID 76xQUARDn]
17. Liu L, Wang Z, Li P. Methods for analysis of medical devices use-related risk: comparison, improvement and application. Indust Engineer Manage 2011;16(6):133-138 [FREE Full text]
18. Masci P, Zhang Y, Jones P, Campos JC. Hazard Analysis Method for Systematic Identification of Safety Requirements for User Interface Software in Medical Devices. 2017 Presented at: 15th International Conference on Software Engineering

and Formal Methods (SEFM 2017); September 4-8, 2017; Trento, Italy URL: http://www.tinyurl.com/ybqe4rpk [doi: 10.1007/978-3-319-66197-1_18]
19. Harrison MD, Masci P, Campos JC, Curzon P. Verification of User Interface Software: The Example of Use-Related Safety Requirements and Programmable Medical Devices. In: IEEE Transactions on Human-Machine Systems. Gainesville, FL, USA: IEEE; Dec 2017:834-846.
20. Paolo M, Paul C, Harold T. Early identification of software causes of use-related hazards in medical devices. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare. 2015 Oct Presented at: MOBIHEALTH'15; October 14-16, 2015; London, UK. [doi: 10.4108/eai.14-10-2015.2261720]
21. Baidu. Object, energy and information are three major elements of physical world URL: https://tieba.baidu.com/p/4210902827 [accessed 2018-03-10] [WebCite Cache ID 6xoXhKTmE]
22. Wen Q, Tang X, Liu X. The study on classification of medical device adverse events based on epidemiological model of injury. Chinese J Pharmacovigilance 2011;8(11):658-661.
23. Dong F, Cheng Y. Active Medical Device Products and Adverse Events Monitoring. Beijing: China Medical Science Press; Sep 1, 2013:66-73.
24. Wang C, Liu Z. An overview of the harm of FDA in the risk management of infusion pumps in the United States. Capital Medicine 2012:4-7.
25. ECRI Institute. 2011 Jan. Top 10 Health Technology Hazards for 2011 URL: https://www.ecri.org/EmailResources/PSO_Monthly_Brief/2011/PSO_Brief_Jan11.pdf [accessed 2019-03-12] [WebCite Cache ID 76owkVliC]

Abbreviations

D: device
E-D: environment-device
FDA: Food and Drug Administration
IQR: interquartile range
MAE: medication administration error
O-D: operator-device
PC: personal computer
P-D: patient-device
PSO: Patient Safety Organization
SHEL: software, hardware, environment, and liveware
TNE: tolerable negative error
TOE: total opportunity for error
U: unknown
UI: user interface

Edited by G Eysenbach; submitted 10.03.18; peer-reviewed by P Curzon, T Poder; comments to author 17.08.18; revised version received 27.02.19; accepted 05.03.19; published 03.05.19.

Please cite as:
Gao X, Wen Q, Duan X, Jin W, Tang X, Zhong L, Xia S, Feng H, Zhong D
A Hazard Analysis of Class I Recalls of Infusion Pumps
JMIR Hum Factors 2019;6(2):e10366
URL: http://humanfactors.jmir.org/2019/2/e10366/
doi:10.2196/10366
PMID:31066695

Original Paper

Usable Mobile App for Community Education on Colorectal Cancer: Development Process and Usability Study

Muhamad Fadhil Mohamad Marzuki[1*], MBBS, MPH; Nor Azwany Yaacob[1*], MD, MCM; Najib Majdi bin Yaacob[2*], MD, MPH, DrPH; Muhammad Radzi Abu Hassan[3*], MBBChBAO, LRCP, MRCP, MMed; Shahrul Bariyah Ahmad[4*], MBBS, MPH

[1]Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kota Bharu, Malaysia

[2]Unit Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, Kota Bharu, Malaysia

[3]Clinical Research Centre, Hospital Sultanah Bahiyah, Alor Setar, Malaysia

[4]Non-Communicable Disease Unit, Kedah State Health Department, Alor Setar, Malaysia

[*]all authors contributed equally

Corresponding Author:
Muhamad Fadhil Mohamad Marzuki, MBBS, MPH
Department of Community Medicine
School of Medical Sciences
Universiti Sains Malaysia
Kubang Kerian
Kota Bharu, 16150
Malaysia
Phone: 60 194743054
Fax: 60 9767 6654
Email: fadhilmarzuki@gmail.com

Abstract

Background: Participation in colorectal cancer screening is still low among Malaysians despite the increasing trend of incidence, with more than half of the new cases being detected in the advanced stages. Knowledge improvement might increase screening participation and thus improve the chances of disease detection. With the advancement of communication technology, people nowadays prefer to read from their mobile phone using a Web browser or mobile apps compared with the traditional printed material. Therefore, health education and promotion should adapt this behavior change in educating the community.

Objective: This study aimed to document the process of designing and developing a mobile app for community education on colorectal cancer and assess the usability of the prototype.

Methods: The nominal group technique (NGT) was used for the content development of the mobile app. NGT involving community educationists and clinicians combined with community representatives as the target users identified relevant health information and communication strategies including features for a user-friendly mobile app. The prototype was developed using framework Ionic 1, based on the Apache Cordova and Angular JS (Google). It was published in the Google Play store. In total, 50 mobile phone users aged 50 years and above and who had never been diagnosed with any type of cancer were invited to download and use the app. They were asked to assess the usability of the app using the validated Malay version of System Usability Scale Questionnaire for the Assessment of Mobile Apps questionnaire. The One-sample *t* test was used to assess the usability score with a cut-off value of 68 for the usable mobile app.

Results: The Colorectal Cancer Awareness Application (*ColorApp*) was successfully developed in the local Malay language. The NGT discussion had suggested 6 main menus in the *ColorApp* prototype, which are Introduction, Sign and Symptoms, Risk Factors, Preventive Measures, Colorectal Cancer Screening Program, and immunochemical fecal occult blood test kit. A total of 2 additional artificial intelligence properties menus were added to allow user-*ColorApp* interaction: Analyze Your Status and *ColorApp* Calculator. The prototype has been published in the Google Play store. The mean usability score was 72 (SD 11.52), which indicates that *ColorApp* is a usable mobile app, and it can be used as a tool for community education on colorectal cancer.

Conclusions: ColorApp mobile app can be used as a user-friendly tool for community education on colorectal cancer.

(JMIR Hum Factors 2019;6(2):e12103) doi:10.2196/12103

KEYWORDS

colorectal cancer; mobile app; development; mHealth

Introduction

Colorectal cancer has become a prominent health problem in the twenty-first century worldwide [1]. It is the third commonest cancer and the fourth leading cause of cancer-related deaths in the world. Its burden is expected to increase by 60% to more than 2.2 million new cases and 1.1 million cancer deaths by 2030 [2]. Malaysia had reported colorectal cancer as the commonest cancer in male and the second commonest cancer in female after breast cancer with an age-standardized rate of 14.6 and 11.1 per 100,000 population, respectively, in 2016 [2].

Colorectal cancer is one of the preventable diseases that are treatable with early detection and treatment. The 5-year survival rate is highly dependent on the stage at diagnosis. Its survival ranges from 95% if detected at stage 1 to 8% if detected at stage IV [3]. In Malaysia, 65% of colorectal cancer was detected at stages III and IV [4], giving rise to lower 5-year relative survival by stage as compared with other developed Asian countries [5]. The late detection might be partly because of the low participation in screening among Malaysians; hence, this called for a more effective strategy to improve in disease knowledge [3,4].

Health education and promotion in Malaysia had utilized a variety of communication strategies including health talk and forums as well as printed material such as pamphlets and health notices. Websites on healthy lifestyles had been created with interactive self-risk-assessment features. However, these might be able to reach those who seek health information either from community health activities or online. With the increasing number of registered mobile phone users, including people in the higher age group in Malaysia [6,7], health outreach thus needs a new strategy. The Interim Review of Malaysian Citizens Reading 2014 had reported that people nowadays prefer to read from their mobile phone using a mobile app or Web browser as compared with printed material [8]. Access to information through mobile phone also seems to be more convenient than attending community activities. Thus, this study encroaches a new strategy in line with the current community behavior by developing a mobile app for a disease of public health importance. This paper aimed to document the process of designing and developing a mobile app for community education on colorectal cancer and report its usability assessment.

Methods

The development of the mobile app prototype was conducted from November 2017 to February 2018. The developmental processes include content development, prototype development, and prototype usability assessment by the intended user. The details of the processes are provided below.

Content Development

The content of the mobile app prototype was developed based on the theory of Health Belief Model (HBM). Content development of the prototype involved a few stages. The first stage was conducting a literature review on colorectal cancer, sign and symptoms, associated factors, screening and prevention, and the features of the mobile app for health promotion and education. This evidence-based information served as important guiding points for the next stage in identifying content relevant to the intended users. For example, participants may not be familiar with the features of a usable mobile app and they may face difficulties in giving their opinion about the features of a user-friendly app. Therefore, researchers provided information on the expected features of a mobile app by the user from published researches.

Nominal group technique (NGT) was applied to explore the following 3 elements:

1. What is the information needed by the intended users?
2. How to ensure that the information is self-explanatory?
3. What are the features needed to make the mobile app a user-friendly app?

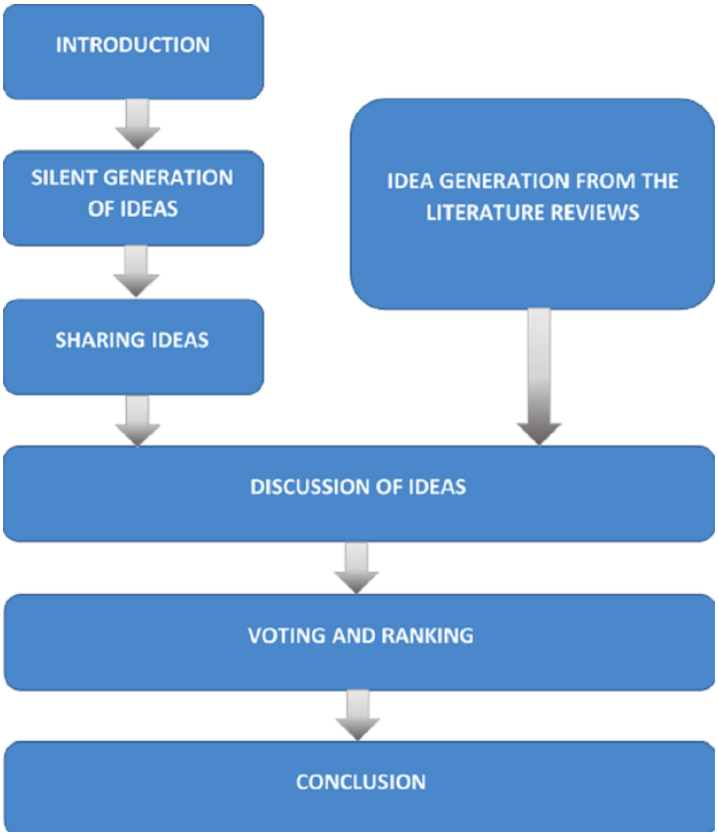
NGT is a structured meeting that aims to provide an orderly procedure for obtaining qualitative information from target groups who are most closely related to the problem area [9]. It is a variation of small group discussion to reach a consensus. NGT gathers information by asking participants to respond to questions posed by a moderator (researcher) and then asking participants to prioritize the ideas or suggestions of all the group members [10]. This technique was developed by Delbecq and Van de Ven and comprises 4 key stages: silent generation, round robin, clarification, and voting (ranking or rating) [11].

The NGT participants consist of the following:

1. one public health physician
2. one gastroenterologist
3. one family medicine physician
4. two medical officers
5. one Assistant Environmental Health Officer (Noncommunicable Disease Control Unit)
6. three individuals from intended users
7. two researchers as moderator

Participants were contacted by phone 2 weeks before the session. They were briefed on the topic and what is the expected outcome from the discussion.

Figure 1. Flow of the process in the Nominal Group Technique.



The session was conducted in the following sequence (Figure 1):

1. Introduction to problem statement and explanation: Moderator briefed the participants on the flow of the session. They were reminded about the expected outcome from the discussion and emphasized on the importance of their contribution to the public benefit.
2. Silent generation of ideas in writing: Each participant was asked to write down their own ideas on what information on colorectal cancer that the community would like to know and what are the mobile app features that will make the apps user friendly and attractive to be used. This session does not allow any discussion and ended after 20 min.
3. Sharing ideas: The participants were invited to share their answers in the round robin manner. Each idea was written on the white board by the moderator. This session does not allow any discussion or argument of the ideas and ended after all participants had shared their answers.
4. Discussion of ideas: The participants were invited to give a verbal explanation about their answers that had been written on the board. They were asked to justify the need of the information and the prototype features to be included in the mobile app. Other participants were encouraged to actively discuss the relevancy of each point.
5. Voting and ranking: Participants were asked to vote their agreed answers based on the discussion. The votes will be tallied to produce the ranking of ideas from highest to lowest based on the research questions.
6. Conclusion: The session was concluded by the moderators. The agreed points were written on the white board.

Prototype Development

The mobile app prototype development was conducted from December 2017 to March 2018. It was developed using the Ionic Framework that is based on the Apache Cordova and Angular JS. A total of 2 researchers who attended training of mobile app development had developed the prototype on the Android platform, which is the main platform for the majority of mobile phones in Malaysia [12]. The NGT findings were used to guide the content as well as features of the prototype. The prototype was then pretested using the same NGT participants for the comprehensibility and clarity as well as technical error. All feedback from the pretests were reviewed and addressed accordingly. The finalized prototype, named *ColorApp* (Colorectal Cancer Awareness Application), sized 9.5 MB was successfully uploaded in early March 2018 onto the Google Developer Console (Google Inc, Mountain View, CA, USA) and published in the Google Play store as a beta option for assessment purposes. It will be released free of charge as a production option for public use as part of the public health contribution once this study is completed.

Prototype Usability Assessment

A cross-sectional study was conducted from February 2018 to March 2018 to assess the usability of the mobile app prototype. It was conducted along with the national-level community empowerment program, *Komuniti Sihat Pembina Negara* (KOSPEN; translates as *Healthy Community, Nation Builder*), in Kota Setar district in Kedah, which is in the northern state of Peninsular Malaysia. KOSPEN is one of the government initiatives for empowering the community for the prevention and control of non-communicable disease in Malaysia. There

are a total of 153 localities in Kota Setar district, covering more than 20,000 people. This area was predetermined because of the distribution of KOSPEN localities in both urban and rural areas and good mobile broadband coverage.

Sample size was calculated using Gpower 3.1.9.2 for mean difference from a constant (1 sample case). The null hypothesis (H_0) was the mean usability score of *ColorApp* is equal to 68 [13] and the alternative hypothesis was 73 (5 unit difference to H_0). These hypotheses were analyzed using the One-sample *t* test. The alpha error was set at 5% and power was set at 80%. The estimated sample size was 33 after considering the detectable difference of 5 units. After an additional 25% anticipated community-based research dropout, the required sample size was 46. As 5 KOSPEN areas were selected, the sample size was rounded to 50. A total of 5 KOSPEN localities were randomly selected. In total, 10 participants aged 50 years and above who use mobile phone with the Android platform and had never been diagnosed with any type of cancer were randomly selected from each locality based on a list provided by the KOSPEN volunteer in the locality to be included in this study. All eligible participants were invited to attend 2 meeting sessions at 2 weeks interval. The first session introduced them on the mobile app. The participants were asked to download and install the mobile app from the Google Play store. They were required to use and interact with the mobile app at their own convenient time within the next 2 weeks. The second session was held after 2 weeks, where all participants were required to answer a validated questionnaire to assess the usability of the mobile app.

Measuring Tool

The usability of the mobile app prototype was assessed using the validated Malay version of *System Usability Scale Questionnaire for the Assessment of Mobile Apps* (in Malay language, it is known as *Skala Kebolehgunaan Aplikasi Mudah Alih*). It is a 10-item questionnaire translated from the original English version, which is the system usability scale (SUS) questionnaire, into local Malay language. This Malay version had been validated to assess the usability of a mobile app [14]. The response score is calculated using the 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The overall score is computed as the summation of all item scores multiplied by 2.5 and can range from 0 to 100. A standard usability score value of 68 was recommended by the original author to indicate good usability of an app [15]. The Cronbach alpha is 0.85. For the purpose of assessment of this mobile app, it was constructed as an online questionnaire using the Google form and the URL link was embedded in the mobile app prototype. Data entry and statistical analysis were made using IBM SPSS version 24.0. The 1-sample *t* test was used to determine whether the mean (SD) usability score of this mobile app is significantly higher than the standard usability score value of 68, with a level of significance (alpha error) less than .05.

This study has been approved by the National Medical Research Registry, Malaysia (NMRR-17-2623-38675 [IIR]), and Human Research Ethics Committee USM, Malaysia (USM/JEPeM/17110601).

Results

Nominal Group Technique Outputs

All participants have agreed that the mobile app prototype must cover the following topics:

1. Introduction to colorectal cancer
2. Sign and symptoms
3. Risk factors
4. Prevention
5. Colorectal cancer screening program

Malay language was chosen to be the main language as it is the national language, with consideration to add other languages in the future. Colorectal cancer was agreed to be translated into the Malay language as *Kanser Kolorektal* instead of the general term of bowel cancer, which is *Kanser Usus* in Malay. This terminology of *Kanser Kolorektal* was agreed to be used to introduce this disease to the community as it is more precise in referring to the disease. The introduction on colorectal was agreed to be of importance to introduce the terminology. The intended user group representative had agreed that the terminology was acceptable and understandable to the community after reading the introduction section.

The features of the mobile app prototype that were agreed upon are simple to operate, using easily understandable language, point-form information, infographic design, include a video, and interactive.

Mobile App Prototype

The mobile app prototype was developed for the Android platform. It is called *ColorApp*, the abbreviation of Colorectal Cancer Awareness Application. Figures 2 and 3 show the screenshots of the main content of *ColorApp*.

Section *Do You Know* (Figure 2, panel a) is the first section that will be seen by the user when they open *ColorApp*. It highlighted the important facts on colorectal cancer that the user needs to know. The user also can watch a short video on colorectal cancer. This is the only section wherein the user is required to have internet access. Other sections can be accessed using a drop-down menu (Figure 2, panel b). The user can click any of the menu options to go to the desired section. Figure 2, panel c, shows the section *Introduction*. The user will be introduced to colorectal cancer. Information was delivered in a point form to ease user's reading and understanding. Pictures of colorectal cancer can be accessed by swiping to the right or left. The section *Sign & Symptoms* listed signs and symptoms of colorectal cancer that can be identified by a layman in a pictorial form (Figure 2, panel d). The user can learn about the risk factors of colorectal cancer through the section *Risk Factor* (Figure 2, panel e). Section *Analyse Your Status* is part of the artificial intelligence features of *ColorApp* that enable the user to enter their particulars, such as age, family history, sex, race, smoking status, and diabetes status (Figure 2, panel f). Once the user clicks on the button *Analyse Status*, *ColorApp* will assess user risk and advise the user on the need for colorectal cancer screening (Figure 3, panel a). Figure 3, panel b, shows the prevention of colorectal cancer. This section enables the user to directly call MQuit Tobacco Quitline for stop smoking

service. Screening for colorectal cancer is introduced to the user via the section *Screening Program* (Figure 3, panel c). This section gives information on the criteria for screening, when and where the screening is available. This section is followed by *immunochemical fecal occult blood test kit* step-by-step instructions (Figure 3, panel d). The section *ColorApp Calculator* is another interactive section whereby the user can enter their age, height, weight, and sex to calculate their body mass index and ideal body weight (Figure 3, panels e and f). This section also provides the recommended level for blood pressure, sugar, and cholesterol. The last section is *About ColorApp* (Figure 3, panel g). This section gives information about this mobile app, email, and contact number for further information. The user also can share this mobile app via social media such as Facebook, WhatsApp, and Twitter with their family and friends.

Figure 2. Screenshots of the main content of ColorApp: (a) "Do You Know" page; (b); drop-down menu (c) "Introduction" page; (d) "Sign & Symptoms" page; (e) "Risk Factors" page; (f) "Analyse Your Status" page.

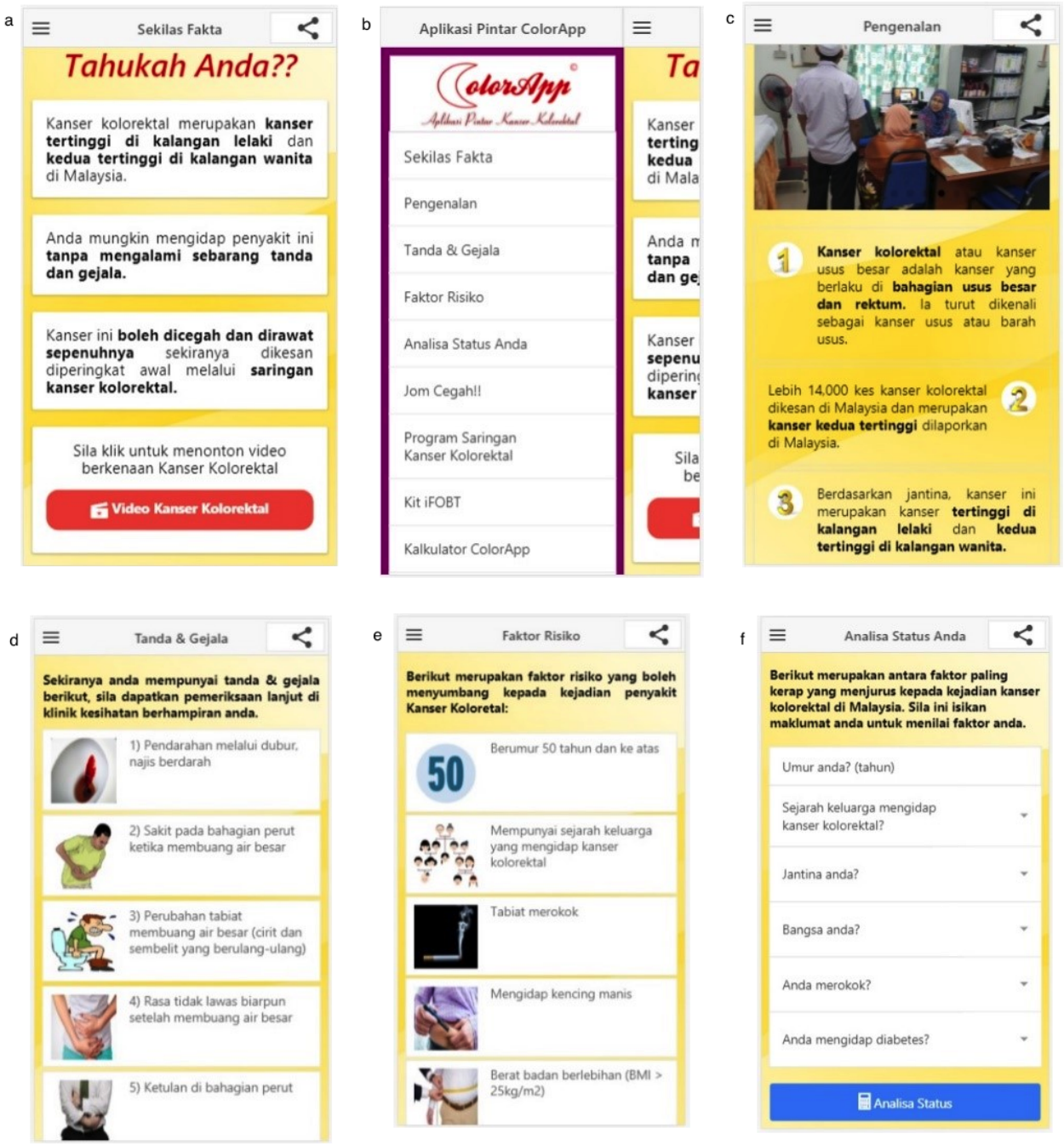


Figure 3. Screenshots of the main content of ColorApp: (a) "Result" page; (b) "Prevention of Colorectal Cancer" page; (c) "Screening Program" page; (d) "iFOBT Kit" page; (e) "ColorApp Calculator" page; (f) "Health Information" page; (g) "About ColorApp" page. iFOBT: immunochemical fecal occult blood test.

Usability

In total, 50 participants were involved in the usability assessment of *ColorApp*. Table 1 shows the sociodemographic characteristics of the participants in this study.

The usability score for *ColorApp* prototype usability showed a mean score difference of 4.9 (P=.004; 95% CI 1.626-8.174), totalling 72.9 (SD 11.52), higher than 68 on the Skala Kebolehgunaan Aplikasi Mudah Alih score which is the minimum cutoff point for a usable system. This indicates good usability of *ColorApp* as a mobile app, and, hence, a usable tool for colorectal cancer community education.

Table 1. Sociodemographic characteristics of participants (N=50).

Characteristics	Statistics
Age (years), mean (SD)	56.0 (5.69)
Sex, n (%)	
Male	25 (50)
Female	25 (50)
Education, n (%)	
Primary	10 (20)
Secondary	33 (66)
Tertiary	7 (14)
Occupation, n (%)	
Unemployed	2 (4)
Self-employed	12 (24)
Retired	1 (2)
Clerical work	32 (64)
Professional	3 (6)

Discussion

Principal Findings

The theory of HBM, which was adopted to develop the content of *ColorApp* prototype, addresses the perceived susceptibility, benefit, and health-seeking behavior of the user. According to HBM, the improvement of health-seeking behavior toward colorectal cancer prevention is related to the perception of an individual that he/she is susceptible to suffer from colorectal cancer, severity of colorectal cancer, and the benefit as well as barrier of preventive action including performing the colorectal cancer screening at the health clinic [16]. All these elements were addressed by the information in textual information, graphic presentation, and video. The interactive page analyzes the user's risk for colorectal cancer and provides advice of action. This will further enhance the perceived susceptibility that the user might have the risk of getting the disease. Combination of various methods of knowledge dissemination will facilitate the knowledge transfer to the user.

The content of *ColorApp* was developed through NGT. NGT is a simplified semiqualitative method to look into preferences of the intended group [17]. NGT allows everyone in the group to contribute to the discussion, and with a good moderator, the likelihood of 1 person dominating the group process can be avoided. NGT also allows identification of important issues and prioritization for content development. Development of health education materials is very challenging, especially when the intended population is from outside the medical field [18]. The community has different levels of health literacy and might face difficulties in understanding certain terms or sentences, especially the health-related terminologies. Thus, the terminology of colorectal was discussed in great detail so as to decide whether to generalize it as bowel cancer or use the specific term *colorectal*. As the introduction had included the simple explanation of what is colorectal, the term colorectal was agreed to be used. Limited health literacy can limit users' understanding, and in a worse scenario, it can pose a risk to them from misunderstanding [19]. The comprehensibility and consideration of health literacy had been addressed well with the involvement of representatives from intended users to ensure that the content of *ColorApp* is delivered clearly, particularly the actionable messages.

People aged 50 years and above are those who are at risk of getting colorectal cancer. This mobile app prototype, thus, is aimed to be used mainly by this group of people. Therefore, several issues that might be faced by the elderly when they use it must be considered. This included issues such as minimalist design to prevent cognitive overload in elderly population, large icons size that is easy to interpret for function and interaction logic, and simple navigation structure (such as back, forward buttons, or menu buttons) [20]. Although the main intended group includes people aged 50 years and above, this mobile app was designed to also be able to attract younger group of users through the combination of different media such as images and video as well as the ability of it being shared via social media with their family and friends, as shown during the NGT [21].

The *ColorApp* was developed as a tool for community education for colorectal cancer prevention. Thus, the usability of the prototype is therefore playing an important role in determining its effectiveness to improve health knowledge and awareness of the user. Usability is defined as the extent to which a product can be used by intended users to achieve specific goals effectively and efficiently as well as providing user satisfaction in a specified context of use [22]. The usability score, which is higher than the minimum acceptable score for a usable mobile app, indicates that *ColorApp* is user friendly for the intended user.

Strength and Limitation

The development of *ColorApp* for community education on colorectal cancer will become the new method in disseminating the information about diseases using a mobile app. The cost of

developing it is relatively cheaper, compared with producing printed materials. It is also easily updated when there are changes in information or management protocols. The features of *ColorApp* that are sharable through social media will help in disseminating this mobile app in the community. Furthermore, the usage of mobile apps is the current trend in information search that should be used to educate the public and improve their knowledge. This app had been registered under the intellectual property of Universiti Sains Malaysia with co-ownership with the Ministry of Health Malaysia and is already available in the Google Play store for all Malay speaking communities that use Android. The term and condition of the Google Play store will be applied with the use of this app. The limitation of this study is that the mobile app prototype is only being developed for the Android platform because of time and logistic reason.

Future Recommendation

In the future, *ColorApp* should be available in other platforms also, especially iOS. The language option also should be made available in English, Chinese, and Tamil as these languages are also used frequently in multiethnic communities in Malaysia. The use of various languages will benefit more users, especially those who prefer certain languages. This will enhance knowledge transfer and improve the user's understanding. Future research also should look into the effectiveness of mobile apps usage in improving the knowledge on colorectal cancer and the attitude and practice toward prevention of this disease including the screening uptake.

Conclusions

In conclusion, the newly developed *ColorApp*, a mobile app for community education on colorectal cancer prevention, which had been developed through NGT and usability process, is a potentially useful tool in the modern technology era. It is available in the Google Play store for free and can be shared and distributed among the community. This study will be extended with the collaboration of the Ministry of Health Malaysia and Clinical Research Center, Hospital Sultanah Bahiyah, Alor Setar, Kedah, to assess the effectiveness of *ColorApp* as a community education and promotion tool in improving user knowledge on colorectal cancer and attitude toward colorectal cancer screening program.

Acknowledgments

The authors would like to express their gratitude to the National Medical Research Registry, Ministry of Health Malaysia; Human Research Ethics Committee, Universiti Sains Malaysia; State Director of Health, Kedah State of Health Department, Kedah Darul Aman; and KOSPEN Community in Kota Setar and Kuala Muda Kedah.

Conflicts of Interest

None declared.

References

1. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. Nat Rev Dis Primers 2015 Nov 05;1:15065 [FREE Full text] [doi: 10.1038/nrdp.2015.65] [Medline: 27189416]
2. Azizah AM, Nor Saleha IT, Noor Hashimah A, Asmah ZA, Mastulu W. National Cancer Institute Ministry of Health Malaysia. Malaysia: National Cancer Institute, Ministry of Health; 2016. Malaysian National Cancer Registry Report URL: https://drive.google.com/file/d/1pCsJIw0ysFccCMFA0XqzfbYu1OFDRbO_/view?usp=sharing [accessed 2019-03-05] [WebCite Cache ID 76clJK0R6]
3. Harmy MY, Norwati D, Noor NM, Amry AR. Knowledge and attitude of colorectal cancer screening among moderate risk patients in West Malaysia. Asian Pac J Cancer Prev 2011;12(8):1957-1960 [FREE Full text] [Medline: 22292632]
4. Koo J, Leong RW, Ching J, Yeoh KG, Wu DC, Murdani A, Asia Pacific Working Group in Colorectal Cancer. Knowledge of, attitudes toward, and barriers to participation of colorectal cancer screening tests in the Asia-Pacific region: a multicenter study. Gastrointest Endosc 2012 Jul;76(1):126-135. [doi: 10.1016/j.gie.2012.03.168] [Medline: 22726471]
5. Veettil SK, Lim KG, Chaiyakunapruk N, Ching SM, Abu Hassan MR. Colorectal cancer in Malaysia: its burden and implications for a multiethnic country. Asian J Surg 2017 Nov;40(6):481-489 [FREE Full text] [doi: 10.1016/j.asjsur.2016.07.005] [Medline: 27492344]
6. Statista. 2017. Number of smartphone users in Malaysia from 2015 to 2022 (in millions)* URL: https://www.statista.com/statistics/494587/smartphone-users-in-malaysia/ [accessed 2019-03-05] [WebCite Cache ID 76cm0RJoY]
7. Malaysian Communications and Multimedia Commission. Selangor, Malaysia; 2017 Oct 16. Internet Users Survey 2016 URL: https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/IUS2016.pdf [accessed 2019-03-06] [WebCite Cache ID 76fXOyMUx]
8. Yusop SH. Berita Harian Onlina. 2017 Jan 15. [Trend of people's reading achieves the target of developed countries] URL: https://www.bharian.com.my/node/234637 [accessed 2019-03-05] [WebCite Cache ID 76cmlBOoS]
9. Yaacob SS. An Interventional Study on Leptospirosis Among Army Personnel in Kelantan: Seroprevalance and Effect on Health Education on Knowledge, Attitude, and Practice. Kelantan, Malaysia: Universiti Sains Malaysia; 2013:88.

10. Centers for Disease Control and Prevention. 2006 Nov. Gaining Consensus Among Stakeholders Through the Nominal Group Technique URL: https://www.cdc.gov/healthyyouth/evaluation/pdf/brief7.pdf [accessed 2019-03-04] [WebCite Cache ID 76cnDClMx]
11. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Jun;38(3):655-662 [FREE Full text] [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
12. Statista. 2017. Number of available applications in the Google Play Store from December 2009 to June 2017 URL: https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/ [accessed 2019-03-04] [WebCite Cache ID 76coIs9JA]
13. Zahra S, Santoso HB. An Indonesian adaptation of the System Usability Scale (SUS). 2016 Oct 15 Presented at: International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2016; October 15-16; Malang, Indonesia URL: https://ieeexplore.ieee.org/document/7872776 [doi: 10.1109/ICACSIS.2016.7872776]
14. Mohamad Marzuki MF, Yaacob NA, Yaacob NM. Translation, cross-cultural adaptation, and validation of the Malay version of the system usability scale questionnaire for the assessment of mobile apps. JMIR Hum Factors 2018 May 14;5(2):e10308 [FREE Full text] [doi: 10.2196/10308] [Medline: 29759955]
15. Sauro J. MeasuringU. 2011. Measuring Usability with the System Usability Scale (SUS) URL: https://measuringu.com/sus/ [accessed 2019-03-04] [WebCite Cache ID 76cngBCbk]
16. Moattar M, Roozitalab M, Gholamzadeh S, Firoozi FS, Zare N. Practical application of health belief model to enhance the uptake of colorectal cancer screening. J Community Med Health Educ 2014 Jul 3;4(4):297 [FREE Full text] [doi: 10.4172/2161-0711.1000297]
17. McMillan SS, Kelly F, Sav A, Kendall E, King MA, Whitty JA, et al. Using the nominal group technique: how to analyse across multiple groups. Health Serv Outcomes Res Method 2014 Jul 25;14(3):92-108. [doi: 10.1007/s10742-014-0121-1]
18. Sentell T, Dela Cruz MR, Heo H, Braun KL. Health literacy, health communication challenges, and cancer screening among rural native Hawaiian and Filipino women. J Cancer Educ 2013 Jun;28(2):325-334 [FREE Full text] [doi: 10.1007/s13187-013-0471-3] [Medline: 23536194]
19. Graham S, Brookey J. Do patients understand? Perm J 2008;12(3):67-69 [FREE Full text] [Medline: 21331214]
20. AppsUsability. 2012 Jun 20. Mobile Apps for Senior Citizen Population URL: http://appsusability.com/2012/06/20/elder/ [accessed 2019-03-04] [WebCite Cache ID 76cnwHiUm]
21. Chan A, Kow R, Cheng JK. Adolescents' perceptions on smartphone applications (Apps) for health management. J Mobile Technol Med 2017 Aug 8;6(2):47-55 [FREE Full text] [doi: 10.7309/jmtm.6.2.6]
22. International Organization for Standardization. 2018 Mar. ISO 9241-11:1998 URL: https://www.iso.org/obp/ui/#iso:std:iso:ts:20282:-2:ed-2:v1:en [accessed 2019-03-04] [WebCite Cache ID 76coGa6LC]

Abbreviations

ColorApp: Colorectal Cancer Awareness Application
HBM: Health Belief Model
H_0: null hypothesis
KOSPEN: Komuniti Sihat Pembina Negara
NGT: nominal group technique
SUS: system usability scale

Edited by G Eysenbach; submitted 08.09.18; peer-reviewed by S Mohd Hairon, W Leclercq; comments to author 27.12.18; revised version received 16.01.19; accepted 17.02.19; published 16.04.19.

Please cite as:
Mohamad Marzuki MF, Yaacob NA, bin Yaacob NM, Abu Hassan MR, Ahmad SB
Usable Mobile App for Community Education on Colorectal Cancer: Development Process and Usability Study
JMIR Hum Factors 2019;6(2):e12103
URL: http://humanfactors.jmir.org/2019/2/e12103/
doi:10.2196/12103
PMID:30990454

Original Paper

Patients' Experiences of Using a Consumer mHealth App for Self-Management of Heart Failure: Mixed-Methods Study

Leanna Sarah Woods[1,2], BN (Hons), Grad Cert Nurs; Jed Duff[3], BN, MN, PhD; Erin Roehrer[4], BCom, BIS (Hons), PhD; Kim Walker[1], BAppSci, PhD; Elizabeth Cummings[1,5], BA, BIS (Hons), PhD

[1]School of Nursing, University of Tasmania, Darlinghurst, Australia

[2]St Vincent's Private Hospital Sydney, Darlinghurst, Australia

[3]School of Nursing and Midwifery, University of Newcastle, Newcastle, Australia

[4]School of Technology, Environments and Design, University of Tasmania, Hobart, Australia

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

Corresponding Author:
Leanna Sarah Woods, BN (Hons), Grad Cert Nurs
School of Nursing
University of Tasmania
1 Leichhardt Street
Darlinghurst, 2010
Australia
Phone: 61 0285727930
Email: leannaj@utas.edu.au

Abstract

Background: To support the self-management of heart failure, a team of hospital clinicians, patients, and family caregivers have co-designed the consumer mobile health app, *Care4myHeart*.

Objective: This research aimed to determine patient experiences of using the app to self-manage heart failure.

Methods: Patients with heart failure used the app for 14 days on their own smart device in a home setting, following which a mixed-methods evaluation was performed. Eight patients were recruited, of whom six completed the Mobile Application Rating Scale and attended an interview.

Results: The overall app quality score was "acceptable" with 3.53 of 5 points, with the aesthetics (3.83/5) and information (3.78/5) subscales scoring the highest. The lowest mean score was in the app-specific subscale representing the perceived impact on health behavior change (2.53/5). Frequently used features were weight and fluid restriction tracking, with graphical representation of data particularly beneficial for improved self-awareness and ongoing learning. The use of technology for self-management will fundamentally differ from current practices and require a change in daily routines. However, app use was correlated with potential utility for daily management of illness with benefits of accurate recording and review of personal health data and as a communication tool for doctors to assist with care planning, as all medical information is available in one place. Technical considerations included participants' attitudes toward technology, functionality and data entry issues, and relatively minor suggested changes.

Conclusions: The findings from this usability study suggest that a significant barrier to adoption is the lack of integration of technology into everyday life in the context of already established disease self-management routines. Future studies should explore the barriers to adoption and sustainability of consumer mobile health interventions for chronic conditions, particularly whether introducing such apps is more beneficial at the commencement of a self-management regimen.

(JMIR Hum Factors 2019;6(2):e13009) doi:10.2196/13009

KEYWORDS

heart failure; mobile health (mHealth); mobile apps; usability study; Mobile Application Rating Scale; patient experience; self-management; mobile phone

Introduction

Heart failure affects at least 26 million people worldwide [1], including more than 1 million Australians [2], and its prevalence is expected to rise [1]. This complex, highly symptomatic syndrome is associated with high health care costs, high readmission rates, and poor clinical outcomes [3]. Targets to improve functional outcomes, psychosocial outcomes, burden of care, and survival of patients with heart failure have resulted in a call for safe, person-centered, evidence-based action [3]. It is especially necessary to ensure equity of care for all patients through the efficient use of resources as well as support to empower patients and caregivers in long-term care [4].

Self-management support, specifically for nonpharmacological requirements, is critical to the effective management of heart failure [2] and is often delivered through educational measures [3,5,6]. Appropriate self-management of heart failure involves daily weight monitoring, fluid restriction, dietary modifications, and exercise alongside regular monitoring and follow-up [2]. In the home setting, recording and recognizing changes such as increased weight, fluid retention, and worsening symptoms, which are indicative of worsening heart failure, can allow patients to get help early [6]. However, challenges with translating guidelines into practice put patients at risk of suboptimal care [2], with the complexity of self-management of heart failure contributing to poor adherence [7].

Rapid improvements in computing capability paired with the popularity of mobile phones in our communities provide more opportunities in health care delivery [7]. Due to this potential, mobile health (mHealth) interventions for heart failure continue to expand; however, this expansion is accompanied by challenges in technology adoption. Reliability of equipment [8], limited technical support [8], cognitive impairment [9], and variable interest in self-recording of health measurements [9] are a few factors affecting use in this patient population. Older people, who have a prevalence of heart failure three times greater than that of the general population [10], have variable levels of willingness to adopt technology [9]. They may lack confidence in their knowledge of heart failure and rely on informal and formal caregivers for guidance [9]. Perceived usefulness and ease of use are considered the most important factors for mHealth adoption [11]. This poses specific challenges when designing interventions aimed to engage patients in self-management of heart failure and highlights the importance of using patient perceptions in newly developed interventions. Further, in a recent review, of the 34 consumer apps targeting heart failure on the commercial app stores, only 3 were evaluated in peer-reviewed articles [12], indicating the importance of disseminating research findings to advance consumer mHealth.

This study is part of a larger research program where *Care4myHeart*, an mHealth app for self-management of heart failure was developed in our hospital by a team of clinicians, patients, and family caregivers. The diverse group of stakeholders collaborated to design an app that was relevant and useful to target users and consistent with the evidence-based heart failure guidelines. The aim of this paper was to explore patients' experiences of and feedback after using the app.

Specific research questions were as follows:

1. What were the patients' experiences of using the *Care4myHeart* app?
2. What is the perceived impact of the app on self-management of heart failure?

Methods

A 14-day usability study was performed using a mixed-methods evaluation to determine patient experiences of using the mHealth app for self-management of heart failure.

Participants

Self-selecting participants were recruited from cardiac inpatient units at a metropolitan private hospital in Sydney, Australia, via posters and flyers located in common patient areas. Medical and nursing staff members were informed of the research and referred patients who voiced their interest in participating. We included English-speaking individuals with heart failure who were not highly dependent on medical care, resided at home, were able to provide feedback, and owned a smart device capable of housing the app. Participants were excluded if they were involved in the co-design of the app, were cognitively impaired, or were otherwise unable to use the app. We aimed for a sample size of 8-10 participants, because up to 80% of usability problems can be identified by this number of users [13].

Intervention

Details of the co-design process of the mHealth app are reported elsewhere [14-17]. The final design of the self-management app has three main sections: *Home screen*, *My Plan*, and *Health Management*. The *Home screen* provides a shortcut to the priority *My Plan* icons based on patient goals, and a reminder summary. The *My Plan* section includes nine important components of self-management of heart failure: medications, symptoms, exercise, weight, fluid, well-being, diet, blood pressure and pulse, and future plans. A *Health Management* section contains a medical documentation repository, appointment calendar, and health care professional contact details. The app provides the opportunity to collect, track, and evaluate patient-entered data. Reminders, alerts, infographics, videos, health professional advice, and information pages throughout the app aim to guide patients to manage their heart failure. Sample user interfaces demonstrating the home, weight, and fluid restriction screens are presented in Figures 1, 2, and 3, respectively.

Figure 1. Sample home screen.

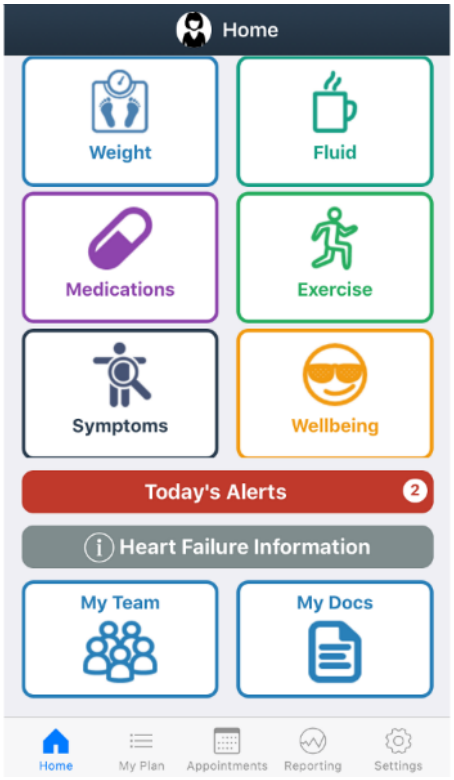
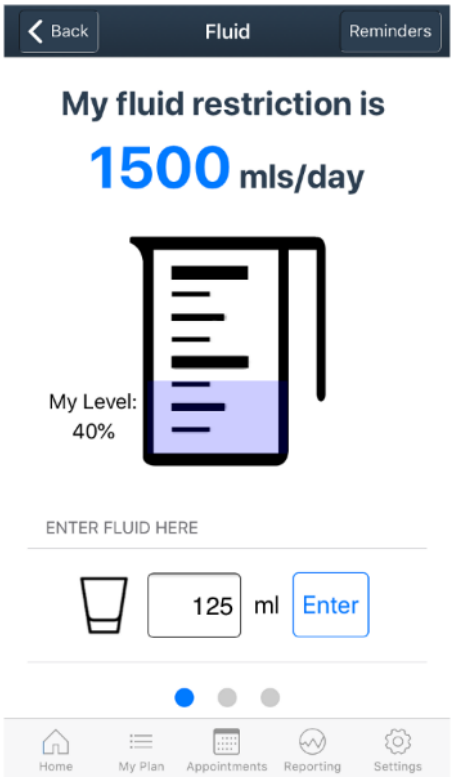


Figure 2. Sample weight screen.

Figure 3. Sample fluid restriction screen.



Study Procedures

The *Care4myHeart* app was downloaded to patients' own iOS or android smartphone or tablet device after procedures were explained and patient consent was obtained. A researcher spent 10-30 minutes providing an overview of the app interface, assisted with completing the personalized settings (dry weight, daily fluid restriction volume, daily step count aim, physical activity goals, and reminders), and determined self-management priorities based on patient preferences. Participants were asked to use the app as frequently as required to assess its usability, aiming for at least daily use over a 2-week period. Participants were encouraged to contact the research team by phone or email if they encountered problems or had questions throughout the study. For quality and safety reasons, participants were instructed to continue with their regular care regime in collaboration with their health care providers. Ethical approval for this study was obtained from the University of Tasmania and St Vincent's Private Hospital Sydney.

Data Collection

As soon as practically manageable after the completion of a 14-day period, participants reported their experience of using both qualitative and quantitative methods.

First, participants were asked to complete the Mobile Application Rating Scale (MARS) [18] either electronically (sent via email) or on paper (sent by post or completed in person during the interview). The 23-item MARS is a multidimensional measure of the four objective app quality indicators: engagement, functionality, esthetics, and information (which together form the overall app quality score). In addition, it includes a subjective quality subscale [18]. As *Care4myHeart* was not available in the app stores during the time of the study, we modified the MARS to 19-items, excluding four items because they were not applicable: accuracy of app description (item 13), goals (item 14), credibility (item 18), and evidence base (item 19). These items were removed from the mean score calculation as per the guidelines [18]. A supplementary, modifiable "app-specific" section assessed the perceived impact of the app on users' target health behaviors [18], in this case, improved heart failure self-management. MARS items are scored on a 5-point Likert scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent) [18]. The version used for this study is provided in Multimedia Appendix 1.

Second, participants were asked to attend an interview on the hospital campus or via phone, depending on patients' preference. A semistructured interview schedule included questions such as "What worked well and what could be improved?" "What functions did you use and why?" and "Would this application impact the way you look after your health?" Participants were given the opportunity to share experiences, communicate thoughts, and voice perspectives through open-ended and probing questions. App use was self-reported by participants themselves, as no usage data were collected in this study. Data were collected in June and July 2018.

Data Analysis

Data were de-identified and treated confidentially. MARS data were managed in the database software program Excel (Microsoft Corp, Redmond, WA), with mean scores produced by calculation of participant subscale scores. Interviews were transcribed verbatim and thematically analyzed using Braun and Clarke's process [19]. The process involved familiarization of the data through re-reading transcripts (Step 1), generation of initial codes and writing them directly on the transcript segments considered interesting or meaningful to the analyst

(Step 2), organization of codes into potential themes (Step 3), review of themes through checking and generating a thematic "map" (Step 4), generation of clear definitions and names for each theme (Step 5), and production of the report with compelling examples through a final analysis (Step 6) [19]. Data analysis in Steps 1 and 2 was conducted by the lead author (LW). Steps 3 to 6 were performed visually and collaboratively, with the themes confirmed by group discussion with the coauthors.

Results

Participant Characteristics

Eight participants consented and commenced the usability study. All participants were male (n=8), most lived with a spouse/partner (n=7) and were currently employed (n=5), and more than half resided in a rural location outside the metropolitan area (n=5). The average age of participants was 69 years (range: 61-84 years).

One participant discontinued the study after reporting technical challenges with a software update that occurred during the 14-day period. A second participant died prior to the final interview and collection of the MARS. Six of the eight participants completed the study with the survey and interview. The interview length ranged from 18 to 29 minutes.

Mobile Application Rating Scale App Quality Scores

Table 1 presents the four subscale scores (engagement, functionality, esthetics, and information), which make up the overall quality score, as well as the subjective quality score (representing satisfaction) and app-specific score (representing behavior change).

The overall app quality score was 3.53 of 5. Of the four subscales, the highest scores were for esthetics (3.83) and information (3.78), followed by engagement (3.37) and functionality (3.33); all scores were above the minimum acceptability score of 3.0. The highest-scoring individual items were layout (4.17), visual information (4.17), interest (3.83), and quality of information (3.80). The lowest scores per item were for performance (2.67), customization (3.00), and interactivity (3.00).

The subjective quality subscale representing app satisfaction scores showed an average of 3.29 of 5. Most participants would use the app more than 50 times in a 12-month period (n=7) and recommend the app to people who might benefit from it (n=4), but would not pay for the app (n=4). The mean star rating, comparable to the star rating on the app stores, was 3.33.

The lowest mean score was in the app-specific section representing the perceived impact of using the app on health behavior change (2.53). The app may have some impact on increased awareness regarding self-management of heart failure (3.17) but was rated "poor" on the perceived impact of the app on attitude, intention to change, help seeking, and overall behavior change (2.33).

Interview Findings

Analysis of interview transcripts resulted in 3 themes and 10 subthemes (Textbox 1).

Theme 1: App Use

Most participants used an android device (smartphones: n=2, tablets: n=2) and two used iPhones. Five participants had both a smartphone and a tablet device. Tablets were kept at home, and smartphones were not necessarily used for internet access. However, those who carry their smartphone in their pocket saw the benefit in data entry throughout the day. iOS users spoke about using their device with greater understanding and confidence than Android users in our sample; the former were also the two youngest participants. Patients self-reported app use for an average of 5-10 minutes once or twice a day on most days during the usability study. The app was used independently without family member involvement. Usage over the 14-day period decreased once users determined what was useful; however, version updates improved technical issues, with usage reportedly increasing after the updates.

Weight, Fluid Restriction, and Step Counter

The weight and fluid restriction sections were most frequently used. The quick speed of recording weight and weight alerts was highlighted as positive features. One participant described how beneficial the fluid recorder was:

> *The most beneficial feature for me at this point in time is the fluid intake...the fluid counter is excellent. I love it, absolutely love it.* [P8]

Fluid volumes were entered either throughout the day or at the end of the day in the fluid restriction section of the app:

> *I wouldn't put in fluid every time I had 100ml of fluid - I put it all in at the end of the day.* [P7]

Some found the app more convenient for self-management of fluid restriction than traditional means of recording fluid volumes because it was portable:

> *Beforehand what I was doing I had a measuring cup...I think the app is more friendly for me to use...I've got that in my pocket, I can always - when I'm out and about - I can make an input on my smartphone and it's just so convenient.* [P8]

To a lesser extent, the step counter within the exercise section was used.

Use of Features

Not all features of the app were used by participants. Participants did not regularly use the symptoms, documents, medication list, and calendar sections, but many saw potential advantages in using these additional features stating, "I didn't use everything but I can see other people could find it very useful" (P1). For example, due to the high frequency of medication changes in patients with heart failure, keeping an updated medication list was perceived as a positive feature. Participants did not use these features during the usability study stating that they "didn't really get a chance to go through it" (P6), and "ah, I had a look but I didn't use any of it functionally" (P7).

Table 1. Mobile Application Rating Scale subscale scores.

Subscale[a] and item	Mean (SD)
Engagement	
Entertainment	3.33 (1.03)
Interest	3.83 (0.75)
Customization	3.00 (0.89)
Interactivity	3.00 (0.89)
Target group	3.67 (0.82)
Subscale mean	3.37 (0.69)
Functionality	
Performance	2.67 (1.63)
Ease of use	3.67 (0.52)
Navigation	3.67 (1.03)
Gestural design	3.33 (0.82)
Subscale mean	3.33 (0.66)
Esthetics	
Layout	4.17 (0.75)
Graphics	3.67 (1.03)
Visual appeal	3.67 (0.82)
Subscale mean	3.83 (0.81)
Information[b]	
Quality of information	3.80 (0.84)
Quantity of information	3.60 (1.52)
Visual information	4.17 (0.41)
Subscale mean	3.78 (0.81)
Overall quality	3.53 (0.63)
Subjective score	
Recommendation	3.50 (1.22)
Use in 12 months	4.67 (0.82)
Pay for the app	1.67 (1.03)
Star rating	3.33 (0.82)
Subscale mean	3.29 (0.70)
App-specific items	
Awareness	3.17 (1.17)
Knowledge	2.67 (0.52)
Attitudes	2.33 (0.82)
Intention to change	2.33 (0.82)
Help seeking	2.33 (0.82)
Behavior change	2.33 (0.82)
Subscale mean	2.53 (0.71)

[a]Mobile Application Rating Scale values range from 1=inadequate to 5=excellent.

[b]The information quality score excluded items 13, 14, 18, and 19 from the Mobile Application Rating Scale.

Textbox 1. Summary of the themes and subthemes from participant interviews.

- App use
  - Weight, fluid restriction, and step counter
  - Use of features
  - Graphs as visual representation of patient data
- Capacity for self-management
  - Established understanding of heart failure and self-management practices
  - App for daily management of illness
  - App as communication tool
- Technical considerations
  - Attitudes toward technology
  - Functionality
  - Data entry
  - Suggested changes

Participants did not watch the instructional exercises videos due to disinterest, personal preference to undertake their own form of exercise, and awareness that they would not continue after a few weeks of watching the same videos. Additional reasons for not using all the features of the app included technical issues and a lack of perceived value for the time required for data entry. One participant commented on why he did not take the time to enter his medications and doctor's contact details into the app:

> *I'm just trying to wait until I get my medications stabilised before I make the inputs...My doctor's names and all of that information I haven't put that in yet but I will over time. It's just – ah – I've I tell you I've been so busy since getting back [home after hospital], just busy busy busy and relaxing after 4 weeks in the hospital.* [P8]

Heart failure information was considered useful for a few patients; however, most participants felt the information was already known to them; one said, "there's no new material for me actually" (P6). Another participant explained how the lack of new information relates to perceived utility of the app:

> *For me it's things I already know...I know I'm big on diet, big on health, so a lot of this information in the app I already know but it just reinforces it...I do enjoy the app but I don't need it.* [P8]

Graphs as a Visual Representation of Patient Data

Visual representation of patient data through graphs was a positive feature of the app, specifically for self-awareness. For daily weight management, graphs were deemed useful, accurate, and relevant and provided feedback to users, as viewing 7-day weight trends heightened self-awareness. A participant explained how the weight trend allowed him to be more "weight aware" (P2), and another appreciated the visual representation of health data specifically:

> *In a graphical sense you see [the weight trend] straight away. And your brain functions on that rather than on just a list of numbers.* [P7]

Self-awareness regarding mobility was deemed beneficial in the exercise section as well. The 7-day step counter graph provided an accurate picture of the mobility status to patients who used the feature:

> *I'm just trying to keep track of how much activity I'm doing, to make sure I'm…keeping moving.* [P1]

Graphical representation of patient data provided learning opportunities. Monitoring the link between fluid intake and fluid congestion can be challenging. However, graphing these data may assist to review previous day's fluid intake and to cross reference this information with fluid congestion symptoms, which may be caused by previous days' nonadherence:

> *[It] appears in your record that you can go back and look and then gives you some sort of positive understanding about what you might have done wrong...your ankles swell up the following morning and you think "ahhhh dopey bugger, I should have bloody been more careful" so and they're lessons we all learn...recognising [I've] gone over [my fluid restriction].* [P7]

Theme 2: Capacity for Self-Management

Participants were unsure how *Care4myHeart* would fit into the way they currently *understand* heart failure and conduct self-management, as using the app for heart failure would require a fundamental change in routine. However, there was potential benefit to heart failure self-management for *daily management of illness* with the benefits of accurately recording and reviewing personal health data, and as a *communication tool* for doctors to assist with care planning, as all medical information is in available one place. These three subthemes are discussed below.

Established Understanding of Heart Failure and Self-Management Practices

Participants found their own way to self-manage their health. Living with the condition for many years, understanding the importance of self-management, and setting goals regarding self-management had contributed to their existing behaviors embedded into daily life. There were many existing self-management strategies: use of a measuring jug on the kitchen bench for fluid intake monitoring, digital calendars, shared household calendar on the back of the pantry door for medical appointments/reminders, liaising with specialist nurses via email, and paper files containing medical documentation.

Participants reported satisfaction with their current health care. Notably, patients reported easy access to health care professionals for regular follow-up, ongoing education/information, and question answering. Participants spoke highly of their current general practitioner, cardiologist, and heart failure nurses:

> *I've got the heart nurse's phone number and mobile number too. She's absolutely fantastic.* [P3]

Participants were aware of and followed a self-management care plan in conjunction with their health care team, knowing their condition is life-limiting. Satisfaction with these current routines was demonstrated:

> *I mean why do I need an app to tell me that ah "do this, do this and this, and you're going to have a better life"? Whereas I get all of this so-called experts, the doctors and all of the information they give you, they tell you the same thing [as the app]...I don't necessarily need an app. Personally, I'm going to do the right thing because I want to live...I know I'm dying. I'm dying as we speak, there's no secrets here but I want to live so I'm going to do the right things.* [P8]

Existing self-management strategies were in a different location or format from the app. Participants compared the convenience of their existing strategies to using the app for self-management. Particularly, participants critiqued the need to "go to various pages on the program" (P3) to view health data, as participants commonly documented information in a notebook or electronic spreadsheet. These existing records have been tailored to the specific requirements considered important by the patients themselves or their health care team. The benefit of these existing daily records was the ability to view their health status at a glance and as a self-management checklist:

> *I can just look at one page and get the whole picture of what's happening...it's all on one page, so I can tick something when I've taken it...I just have a look at [the page] and see that I've done everything that day and basically...well that's the day done, I'm complete.* [P3]

Further, existing strategies were considered easy and time efficient in everyday life, as one participant explained about maintaining his fluid restriction throughout the day using other strategies compared with using the app:

> *I would personally keep going the way I'm going cos of the ease of doing it...[T]he easy things I'd rather just do easy, like the water in the jug...where the app's stuck in my bedroom most of the time. I've gotta go and turn it on, I've gotta go bang, bang, bang, and by the time I've sorta done the water in the jug I've well and truly finished before probably I've even get into the program properly.* [P3]

Although the app may assist in monitoring specific self-management activities like weight or fluid intake, it did not seem to embody the complexity of self-management of heart failure. Participants communicated a good understanding of heart failure (with the exception of one participant who was not familiar with the term "dry weight"). They correctly understood that fluid congestion was variable, fluid intake and diuretic medications are directly linked to fluid status, and regular self-assessment for abdominal/ankle edema was necessary. Understanding these concepts of heart failure involved a more thorough and subjective self-assessment, which was not directly equivalent to the setting's parameters within the design of the app. One participant explained his thought process while conducting a self-assessment, which was a more complex process than simply adhering to a daily fluid restriction:

> *Sometimes I will go over my fluid intake which is 1.2 [litres], sometimes I go over because I'm looking at the way I feel...I'm doing a couple of things. I'm looking at the fluid intake but I'm also looking at my body or seeing the way I feel...I'm looking at how dry I am…I'll just drink a little bit more and not get a doctor review [because] I haven't started to pick up any signs of oedema.* [P8]

App for Daily Management of Illness

The app provided a routine to manage health data like weight. Participants explained that "it generates a discipline to maintain the information" (P2) specifically regarding "the daily management of my fluid balance, it takes a lot of adjustment...to get the balance right" (P1). Entering weight was quicker using the app than the usual format of documenting weight for some proclaiming "this is a quicker way of doing it, like most computers it can store information well" (P2).

Recording health information within the app on a daily basis was considered more accurate than manual measures or memory. One participant explained how he normally relies on memory:

> *I don't record it as such but I check it every couple of days keeping a mental note – I just want to make sure there are no big variations from day to day so that's all I look for [but with the app] it's nice to have that trend, I like it, it gives you a more accurate picture.* [P6]

The health data repository and feedback within the app provided an opportunity to view a person's health status more objectively. For example, accurate recording of health data might help family members seek care appropriately during times of worsening heart failure:

> *If you go into denial stage and don't pay attention to the weight because you don't want to go into hospital*

> *or something, now they can look and see "Ah well that's not right – we should get you to the doctor" so I think it would help.* [P1]

App as Communication Tool

The app was considered a potential tool to communicate with doctors and other health professionals about assisting with care planning. Participants explained how the app could facilitate accurate information sharing:

> *[The app]enables you to communicate with your medical practitioner in a fairly accurate - one would hope - way, about what's been going on and therefore one would hope, if you were the medical practitioner, I suppose it would cause the medical practitioner a better basis of making decisions about your medical care.* [P2]

As a potential communication tool, the app could assist doctors with patient assessment. Participants frequently spoke of the potential to show doctors the graphs representing health-related trends of recent days in a consultation, as "it's quick" (P7), or over the phone, as "If you had it on a phone you could just say [to the doctor] 'Look, I'll send this through to you"' (P7). Another participant agreed with this potential:

> *The concept is good because you can take your tablet along to your doctor and he says "well how have you been?" and you can say "well there you are, there's my weight, there's my blood pressure," so you've got that information available.* [P2]

Having medical information in one place was deemed useful if all relevant data were stored in the app. Digital storage of personal medical records was considered "very powerful and very useful" (P7), as participants saw benefit in having "everything in one place" (P6) and "recorded accurately" (P1). Digital copies of medical information were considered "much easier rather than carrying an actual physical document. Sometimes I forget to take it" (P6). The potential to use the app as a communication tool was deemed especially valuable for new or temporary doctors and during medical emergencies:

> *Just air drop [my current medication list] from your phone to the doctor in casualty or whatever I think's a great, very good idea...I think that would be helpful for a lot of people especially if you come into hospital somewhere hypoxic...unconscious or whatever...or too breathless to talk about it. I've got a very very extensive list of drugs that I'm on, I think it's 35 tablets a day usually, so having that list when I've gotta provide it, makes it much easier.* [P1]

However, no participants reported using the app with members in their health care team during the time of the study. Further, the version used for the usability study was not set up for third-party access.

Theme 3: Technical Considerations

There were technical considerations influencing the experience of using the app, including *attitudes toward technology* and *functionality* and *data entry* issues. These subthemes are reported in the following section alongside the final subtheme—numerous *suggested changes* —to improve the app's design.

Attitudes Toward Technology

Predominantly, participants were not regular users of smart devices for apps or health. Three sample quotes demonstrated minimal interest in using smart devices overall:

> *I'm not a big user of phones, especially mobile phones.* [P8]

> *I don't particularly like turning computers on anyhow, I mean I'd go a fortnight without reading my emails.* [P3]

> *I'm a dinosaur and not used to using texting.* [P7]

Trust was one reason a participant would not use internet banking or purchase products using a credit card (P3). Participants reported using their smart devices for Google calendar, checking the weather forecast, playing games (CandyCrush, solitaire, or crosswords), and internet searches, and only a few used emails. In relation to technology use for health, one participant reported using a health app for self-management of heart failure and another stored his current medication list in the notes section of his smartphone. No participants reported storing medical documents electronically.

Participants believed in the inevitable advancement of technology in the contemporary era, and this was perceived to include the acceptance of health apps like *Care4myHeart* for younger generations. With the everyday use of smartphones, the younger generation "would approach it completely differently" (P7). Another participant explained:

> *I think for really the next generation and computer nerds at the moment you're on a winner there, I really do...As you get the younger ones come through you'll be fine, which will happen just over time.* [P3]

Attitudes toward technology by family members appeared consistent with those of the participants. There were no reports of receiving assistance from family members by using the app:

> *[My wife is] less techno-cradic [sic] than I am. I mean she went from a phone with a touchscreen back to a phone with push buttons on it, that's what she likes.* [P7]

The personal nature of smartphones may impact the divide between family members:

> *[It is] my phone so she didn't really take a closer look.* [P8]

Functionality

Technical challenges were reported to affect usage, which was more prevalent in Android than iOS devices. Issues with downloading the version update on Android caused one participant to discontinue participation in the study. A second participant was unable to download the updated Android version but managed to continue with the original version downloaded at the beginning of the usability study:

> *The whole thing stands still. Still. Still doing nothing...The process of downloading the app is pretty*

> *clearly signposted, I'm not complaining about that, it just didn't work.* [P2]

Technical issues with the Android version also included: lengthy app loading; a blank 7-day weight graph; and the inability to record blood pressure readings, set medication reminders, and use the clock function. Virus-protection interference due to the app being from an unknown source was also reported, regardless of approval of unknown sources in the settings section of the device. The iOS version had less technical issue reports overall but a lengthier multistep initial download process and intermittent screen freezes.

Technical issues were a barrier for ongoing use. Participants commented on the ongoing struggles with the usability:

> *I've persevered with it...but I found I was battling [with the app].* [P7]

> *Whether it's me or whether it's the program or a combination of both I don't know, but that's your problem.* [P3]

The potential benefit of the app versus the technical challenges associated with the app was also reflected:

> *I still think the idea is good and I think it's easy enough to use if it works but I've still got problems with the execution, you know.* [P2]

Interestingly, participants seldom reported technical challenges encountered by the research team during the usability study but raised these issues during the interview.

Data Entry

Navigation and data entry were specifically problematic. Participants reported physical limitations during the operation of the app, saying they have "big clumsy fingers" and their "hands shake a little bit" (P7). Participants experienced time-consuming data entry in the medications section, challenges with using some buttons, and confusion completing or updating the settings.

Strategies to overcome these limitations were evident, as participants had insight into their own ailments:

> *Sometimes I lick the end of my fingers and that might be a factor of fluid, my fluids are very low and I'm quite dry.* [P7]

Awareness of these functional limitations was a factor in participants choosing a tablet device over a smartphone if they owned both: "I've got fat fingers and the phone's got a small keyboard" (P2). Further, the consequences of incorrect data entry in the settings component of the app caused inappropriate alerts. One participant explained an alert associated with incorrect entry of dry weight:

> *It told me horror stories about what I should do in terms of consulting my medical practitioners, when in fact I had simply a [settings] error on the machine.* [P2]

Suggested Changes

Many suggested changes were provided in relation to data entry issues, utility by the heart failure population, and making it more appealing for the user.

There were many usability improvements regarding the data entry challenges experienced. Participants wanted more control over their data: "people are generally pretty honest about the way they deal with their own data" (P7). Participants wanted to clear previously entered or incorrect data, edit previously entered data, and enter retrospective data in case it was missed, causing incomplete weight graphs:

> *If you're out for the day say and you leave your phone at home and you come back and want to add the data the following day, you can't do it, so I think that is definitely a negative.* [P7]

Having an empty data entry screen without predicted or previous amounts was important to avoid confusion during data entries. This was noted for documenting fluid intake and entering daily weight:

> *It comes up with the last weight you put in so you have to delete that before you can actually [put] a revised weight in and I think that's a mistake. I think the window should be clear and you just enter in the data you want to enter.* [P7]

In addition, there were suggestions to improve the applicability to the patient group. These included recording more health data, documenting medication variations more easily, adding a medication checklist function, going over the maximum fluid restriction volume, and adding a free-text general notes page.

Making the user interface more appealing was deemed necessary for engagement with the app. Suggestions included visualization of fluid overflowing out of the fluid jug or turning red in color and more graphical information with an increase to a 14-day trend. Participants explained their wish for a more interesting interface:

> *If you can have some whistles and bells and things like that–it just makes it a little bit more interesting.* [P8]

> *Some screens are very average looking...I think if you could brush it up a little bit and um, make it more appealing some of the screens...would be nice actually.* [P6]

These improvement suggestions would perceivably improve the utility of the app:

> *[To] make notes about day to day things…just like a general notes page. That would be a great idea...That would be the decider for me to use it over the other one [app].* [P1]

Miscellaneous suggested changes included a simpler keyboard, ability to change to horizontal view on the tablet version, and appearance of the logo on more screens.

Discussion

Learning from Failure

This paper presents findings from a usability study conducted with patients using an mHealth app for heart failure. We explored the way the app was used and its perceived impact on self-management of disease. In this context, frequently used features were weight and fluid restriction tracking, and graphical representation of data was particularly beneficial. Using technology for self-management would fundamentally differ from current practices; however, use of the app was correlated with the potential utility for daily condition management and as a communication tool. The overall app quality score, as assessed by the MARS, was slightly higher for *Care4myHeart* (3.53) than an average of the 34 comparable heart failure-support apps on the consumer app stores (3.4) [12]. In its current form, the perceived impact on health behavior change was classified as "poor" in the MARS app-specific subscale. Patient experiences of using various app components highlighted challenges and opportunities for design improvements for the next version of the *Care4myHeart* app. In addition, patient experiences have implications for researchers investigating digital health systems for chronic disease and consumer app designers wishing to incorporate human factors. Many lessons were learned from the usability study and are described below.

Lessons Learned

The following lessons were learned from the evaluation of *Care4myHeart* by patient participants.

Lesson 1: If Technology Is Not Integrated Into Everyday Life, It Is a Significant Barrier to Adoption

Integrating self-management with normal life patterns has been identified as a key enabler of effective self-care in heart failure [20], and participants in this study have well-established daily routines. Clarke et al [20] described how patients with heart failure enlist "cues" in everyday life as routines to facilitate guideline adherence. For example, to integrate self-management activities with the morning routine, patients may place pill boxes on the breakfast table as a visual reminder for medication adherence [20]. Participants in the usability study for *Care4myHeart* reported various cues and, except for a few, reported their ease and desire to continue with the existing routines. Demonstrating this, the use of a measuring jug on the kitchen bench for daily fluid restriction management served three functions: a visual reminder to limit oral fluids, a functional measuring tool, and an accurate visual representation of cumulative fluid intake at any point in the day. This presents a more convenient option for participants whose smart devices were located elsewhere in the house and had a more practical option, given the inability of the technology to measure fluid volumes. Participant reflections in comparing the use of technology in heart failure were consistent with the recent study conducted with older people with heart failure: Nguyen et al [9] found that "Some patients did not find technology to be useful or relevant in their daily activities because they were already comfortable with their routines." Similar reasons likely contributed to the low perceived impact of the app on health behavior change reported in the MARS and indifference to explore all app features, as participants felt the app did not enhance existing self-management. Consequently, introducing the app at the commencement of a self-management regimen may be more beneficial and needs further investigation.

The private nature of smart devices may be a barrier to adoption itself. In this study, no participants reported the involvement of family caregivers regarding the use of the *Care4myHeart* app. Yet, historically, caregivers are frequently involved in heart failure [21] with some patients dependent on their caregivers to make health-related decisions [9]. The gradation of dependency of caregivers for older adults with chronic conditions [22] presents challenges in designing future support interventions [20] when daily health-related activities involve caregivers. The technology risks excluding caregivers unless the design supports their active involvement and the resulting design presents a perceived benefit to the patient *and* caregiver.

Lesson 2: The Biggest Benefit Is the Opportunity for Improved Self-Awareness and Continuous Learning in Heart Failure Management

The timely detection and recognition of and action to subtle changes in symptoms was noted as a key skill for effective self-management of heart failure [20]. According to patient experiences, the self-management app we developed offered possibilities for a more active role in daily recording and reviewing of heart failure-related data. Participants specifically observed a benefit in the graphical representation of their data with the ability to view trends, detect changes representative of worsening heart failure, and take action accordingly. Previous studies have shown that skills in managing heart failure evolve over time and learning from past experiences are helpful in applying effective strategies to daily life [21]. This was particularly evident with patients' experiences using the 7-day weight trend feature. Participants felt it was accurate and timely and provided an objective representation of their health status to watch or act when needed. We believe that the use of mHealth via an app with real-time representation of data trends would strengthen patient empowerment and decision making in self-management.

However, to realize the potential for improved self-awareness and continuous learning, engagement improvements are needed. A recent review, which compared the quality of 34 heart failure support apps on the consumer app stores using the MARS, found the lowest score was for the engagement subscale (2.9/5.0) [12]. This led to a call for further improvements in engagement of mHealth apps for heart failure support. In the context of our study, *Care4myHeart* had an engagement subscale mean of 3.37, which was higher than the average in the review. However, this score still falls short of the "good" range. In this regard, participants conveyed valuable suggestions to improve the interactivity and customization of the app, in addition to suggestions to make the interface more interesting and entertaining. Incorporating the many suggestions provided from (just) six participants in the study may greatly improve the interface for future users. The suggested changes are relatively minor to incorporate in iterations, as they have been in other usability studies [23] achieved through usability studies of similar sample sizes of 5-10 participants [24-26].

Lesson 3: Patients Need a Way to Manage Their Health Information Across the Health Care System

The findings of this research indicate that participants want effective ways to share their data with health care professionals for ongoing care. Participants perceived the app to be effective as a communication tool to share their data in a timely, accurate, and visual manner, so that health care professionals can be armed with all relevant health information contained in one system, especially in an emergency or unfamiliar health care setting, for care planning. Australia is transitioning to an opt-out electronic health record; however, during the usability study period, participants' health information was largely held in silos by individual health providers. Participants reported the safety and quality benefits to record, store, and manage health information in one place, whether it was the *Care4myHeart* app or another assistive technology. These participants' perspectives are mirrored in a recent study investigating experiences using the patient-accessible electronic health record used in Sweden [27]. Over 96% of survey responders had an overall positive perception of the system, reporting the following highest-rated reasons why they felt it important to have access to their health-related information: (1) it makes patients feel informed, (2) it improves communication between medical staff and the patient, (3) it improves the understanding of the patient's condition, and (4) it makes patients feel safe [27].

Condition-specific mHealth apps have limitations for integration to current health information systems across acute care, primary care, and community care. Standalone apps will not reach their potential to aid self-management without integration across health care providers, because, like other chronic conditions, patients with heart failure have concurrent comorbid conditions [1], experience frequent hospitalizations [3], and require a team approach across health care sectors [5]. There is increasing recognition that health services for those living with chronic conditions need to be more integrated, coordinated, and patient focused across the continuum of care [2]; however, mHealth has specific challenges in addition to other service redesign efforts. For example, health system readiness, organizational resistance to change, policy uncertainties, and unclear reimbursement schedules for clinicians have been previously identified as barriers to the successful implementation of mHealth technologies for chronic conditions [22].

Lesson 4: Technical Challenges are a Significant Barrier to Use With Most Patients Unlikely to Persevere

Attitudes toward technology use impacted participants' experiences of using the app. The complex components within the app requiring more navigation and data entry, for example, the medication list feature, were infrequently used. These complex components were more likely to have technical and functional issues, which was an additional deterrent reported by participants with less confidence of using technology. For the few participants who self-reported daily app use, the technical challenges were less of a hindrance, but these participants were more likely to provide specific interface-improvement suggestions.

The findings of this usability study have led to recommendations regarding technology use for usability studies conducted with patients, which may be particularly beneficial to clinician researchers. First, testing and re-testing before allowing patients to use the technology is important to help mitigate frustration of poorly functioning technology, a previously reported fear in older adults with heart failure [9]. Second, avoiding version updates during a usability trial will limit confusion, particularly when the researcher cannot screen share with patients located in rural areas to guide the process. Finally, consider recruiting patients who use apps daily as "early adopters" of mHealth for heart failure because of the variable levels of technology acceptance in this patient population [9]. Our findings were consistent with those of Nguyen and colleagues [9] who found that patients were keen to manage their heart failure and willing to uptake self-management recommendations, but discovered that for some patients, adopting a new technology on top of their daily health routines may be of little benefit. Time and effort were barriers to technology acceptance [9], consistent with the findings from this study, where the ease and convenience of continuing with existing self-care regimens outweighed the technical challenges of learning how to use a new app. This would also account for the seldom reporting of technical difficulties during the study. Participants likely made decisions about their acceptance of the app early in the study period and therefore lacked motivation to troubleshoot technical issues with the research team. We found these barriers to technology use regardless of the participant's keen interest to participate in the research and optimism for technology to assist with their health, noting that the demographic of study participants were older men only.

We tried to minimize technical challenges by using a participatory, co-design approach involving patients in each stage of the development; however, this was not reflected in the study's findings. This challenges the assumptions of the co-design methodology in addressing the needs of target users and improving usability and places further emphasis on the nonhomogenous attitudes of patients with heart failure when considering technology and health.

Recommendations for Future Research

Future research should explore in what formats and contexts technology can positively complement daily self-management activities conducted by patients with heart failure. Importantly, we must incorporate the vital caregiver role in the design of condition-specific mHealth because of their active role in self-management support in the home environment. A more focused understanding of the design considerations to engage users in an interesting and beneficial way is likely necessary for adoption and ongoing use, which will require interdisciplinary collaboration between designers, developers, health care providers, and health care consumers. Third-party access to medical information in the app, especially in an emergency, may be an important design recommendation and should be investigated.

With the limited number of evidence-based mHealth interventions moving past the pilot or feasibility stage [22], future studies should investigate the many barriers to adoption and sustainability. Implementation science of mHealth apps for self-management of chronic conditions as an adjunct to existing

care is an important area for further research, specifically for investigating perspectives of clinicians, health system administrators, and policy makers.

Limitations

Since data collection, the authors are aware of a user version of the MARS called uMARS [28], which would have suited this participant sample more specifically as health care consumers. A limitation of this research is the selection bias of the patients. First, as per the inclusion criteria, all participants owned a smart device. Second, less adherent patients, for whom the app may be most beneficial, are often not willing to participate and may have reported different experiences from this sample. The findings from this study conducted with a small and homogenous sample cannot be generalized to the wider heart failure population; nevertheless, they provide insight for further research on the topic.

Conclusion

A mixed-methods evaluation of patient experiences using an mHealth app for heart failure showed how the app was used and its perceived impact on self-management. Daily self-management habits are established without the use of technology, so patients were unsure how the app would fit in their routines. Nevertheless, participants saw the potential of the app to aid daily condition management, particularly regarding weight and fluid restriction management, and serve as a communication tool for health care professionals involved in their care.

Understanding users' experiences contributes to design improvements for the *Care4myHeart* app, and the lessons learned have implications for researchers and development teams to advance the quality of consumer mHealth apps for chronic conditions. Future studies should investigate the barriers to adoption and sustainability of consumer mHealth interventions, including whether introducing such apps is more beneficial at the commencement of a self-management regimen. Research into how to incorporate the important role of caregivers in the design of technology to support self-management in the home environment is also needed.

Acknowledgments

We are thankful to the patients for their contributions to this research. The research reported in this paper is supported by a St Vincent's Clinic Foundation grant and an Elite Research Scholarship provided by The District Nurses, Hobart, Australia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Modified Mobile Application Rating Scale.

[PDF File (Adobe PDF File), 107KB - humanfactors_v6i2e13009_app1.pdf]

References

1. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. Card Fail Rev 2017 Apr;3(1):7-11 [FREE Full text] [doi: 10.15420/cfr.2016:25:2] [Medline: 28785469]
2. NSW clinical service framework for chronic heart failure. Version 1.2. Chatswood NSW: Agency for Clinical Innovation; 2017:1-60.
3. Heart failure toolkit: A targeted approach to reducing heart failure readmissions. Victoria: National Heart Foundation of Australia; 2016:1-38.
4. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, et al. Heart failure: preventing disease and death worldwide. ESC Heart Fail 2014 Sep;1(1):4-25. [doi: 10.1002/ehf2.12005] [Medline: 28834669]
5. National Heart Foundation of Australia. Multidisciplinary care for people with chronic heart failure: principles and recommendations for best practice. Australia: National Heart Foundation of Australia; 2010:978-971.
6. Living well with heart failure: Information to help you feel better. Australia: National Heart Foundation of Australia; 2016.
7. Cajita MI, Gleason KT, Han H. A Systematic Review of mHealth-Based Heart Failure Interventions. J Cardiovasc Nurs 2016;31(3):E10-E22. [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
8. Alnosayan N, Chatterjee S, Alluhaidan A, Lee E, Houston Feenstra L. Design and Usability of a Heart Failure mHealth System: A Pilot Study. JMIR Hum Factors 2017 Mar 24;4(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.6481] [Medline: 28341615]
9. Nguyen L, Keshavjee K, Archer N, Patterson C, Gwadry-Sridhar F, Demers C. Barriers to technology use among older heart failure individuals in managing their symptoms after hospital discharge. Int J Med Inform 2017 Dec;105:136-142. [doi: 10.1016/j.ijmedinf.2017.06.001] [Medline: 28750907]
10. Sahle BW, Owen AJ, Mutowo MP, Krum H, Reid CM. Prevalence of heart failure in Australia: a systematic review. BMC Cardiovasc Disord 2016 Feb 06;16:32 [FREE Full text] [doi: 10.1186/s12872-016-0208-4] [Medline: 26852410]

11. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to Use mHealth in Older Adults With Heart Failure. J Cardiovasc Nurs 2017 Feb 28. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
12. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
13. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
14. Woods L, Cummings E, Duff J, Walker K. Conceptual Design and Iterative Development of a mHealth App by Clinicians, Patients and Their Families. Stud Health Technol Inform 2018;252:170-175. [Medline: 30040701]
15. Woods L, Cummings E, Duff J, Walker K. The development and use of personas in a user-centred mHealth design project. New York: ACM; 2017 Presented at: 29th Australian Conference on Computer-Human Interaction (OZCHI'17); Nov 28 - Dec 1; Brisbane p. 560-565 URL: https://doi.org/10.1145/3152771.3156186 [doi: 10.1145/3152771.3156186]
16. Woods L, Cummings E, Duff J, Walker K. Design Thinking for mHealth Application Co-Design to Support Heart Failure Self-Management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]
17. Woods L, Cummings E, Duff J, Walker K. Partnering in Digital Health Design: Engaging the Multidisciplinary Team in a Needs Analysis. Stud Health Technol Inform 2018;252:176-181. [Medline: 30040702]
18. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Clark AM, Spaling M, Harkness K, Spiers J, Strachan PH, Thompson DR, et al. Determinants of effective heart failure self-care: a systematic review of patients' and caregivers' perceptions. Heart 2014 May;100(9):716-721. [doi: 10.1136/heartjnl-2013-304852] [Medline: 24548920]
21. Jaarsma T, Cameron J, Riegel B, Stromberg A. Factors Related to Self-Care in Heart Failure Patients According to the Middle-Range Theory of Self-Care of Chronic Illness: a Literature Update. Curr Heart Fail Rep 2017 Apr;14(2):71-77 [FREE Full text] [doi: 10.1007/s11897-017-0324-1] [Medline: 28213768]
22. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, Implementing, and Evaluating Mobile Health Technologies for Managing Chronic Conditions in Older Adults: A Scoping Review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
23. Tatara N, Årsand E, Bratteteig T, Hartvigsen G. Usage and perceptions of a mobile self-management application for people with type 2 diabetes: qualitative study of a five-month trial. Stud Health Technol Inform 2013;192:127-131. [Medline: 23920529]
24. Athilingam P, Osorio RE, Kaplan H, Oliver D, O'neachtain T, Rogal PJ. Embedding Patient Education in Mobile Platform for Patients With Heart Failure: Theory-Based Development and Beta Testing. Comput Inform Nurs 2016 Feb;34(2):92-98. [doi: 10.1097/CIN.0000000000000216] [Medline: 26765655]
25. Klasnja P, Hartzler A, Powell C, Phan G, Pratt W. Health Weaver Mobile: Designing a Mobile Tool for Managing Personal Health Information during Cancer Care. AMIA Annu Symp Proc 2010;2010:392-396 [FREE Full text] [Medline: 21347007]
26. Burnay E, Cruz-Correia R, Jacinto T, Sousa AS, Fonseca J. Challenges of a mobile application for asthma and allergic rhinitis patient enablement-interface and synchronization. Telemed J E Health 2013 Jan;19(1):13-18. [doi: 10.1089/tmj.2012.0020] [Medline: 23215639]
27. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
28. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]

Abbreviations

mHealth: mobile health
MARS: Mobile Application Rating Scale

Edited by A Kushniruk; submitted 03.12.18; peer-reviewed by S Donnelly, M Cajita; comments to author 22.01.19; revised version received 04.03.19; accepted 21.03.19; published 02.05.19.

Please cite as:
Woods LS, Duff J, Roehrer E, Walker K, Cummings E
Patients' Experiences of Using a Consumer mHealth App for Self-Management of Heart Failure: Mixed-Methods Study
JMIR Hum Factors 2019;6(2):e13009
URL: http://humanfactors.jmir.org/2019/2/e13009/
doi:10.2196/13009
PMID:31045504

Original Paper

Communicating Bad News: Insights for the Design of Consumer Health Technologies

Eun Kyoung Choe[1], PhD; Marisa E Duarte[2], PhD; Hyewon Suh[3], MS; Wanda Pratt[4], PhD; Julie A Kientz[3], PhD

[1]College of Information Studies, University of Maryland, College Park, MD, United States

[2]School of Social Transformation, Arizona State University, Tempe, AZ, United States

[3]Human Centered Design and Engineering, University of Washington, Seattle, WA, United States

[4]Information School, University of Washington, Seattle, WA, United States

Corresponding Author:
Julie A Kientz, PhD
Human Centered Design and Engineering
University of Washington
423C Sieg Hall, Box 352315
Seattle, WA, 98195
United States
Phone: 1 2062210614
Email: jkientz@uw.edu

Abstract

Background: As people increasingly receive personal health information through technology, there is increased importance for this information to be communicated with empathy and consideration for the patient's experience of consuming it. Although technology enables people to have more frequent and faster access to their health information, it could also cause unnecessary anxiety, distress, or confusion because of the sensitive and complex nature of the information and its potential to provide information that could be considered bad news.

Objective: The aim of this study was to uncover insights for the design of health information technologies that potentially communicate bad news about health such as the result of a diagnosis, increased risk for a chronic or terminal disease, or overall declining health.

Methods: On the basis of a review of established guidelines for clinicians on communicating bad news, we developed an interview guide and conducted interviews with patients, patients' family members, and clinicians on their experience of delivering and receiving the diagnosis of a serious disease. We then analyzed the data using a thematic analysis to identify overall themes from a perspective of identifying ways to translate these strategies to technology design.

Results: We describe qualitative results combining an analysis of the clinical guidelines for sharing bad health news with patients and interviews on clinicians' specific strategies to communicate bad news and the emotional and informational support that patients and their family members seek. Specific strategies clinicians use included preparing for the patients' visit, anticipating patients' feelings, building a partnership of trust with patients, acknowledging patients' physical and emotional discomfort, setting up a scene where patients can process the information, helping patients build resilience and giving hope, matching the level of information to the patients' level of understanding, communicating face-to-face, if possible, and using nonverbal means. Patient and family member experiences included internal turmoil and emotional distress when receiving bad news and emotional and informational support that patients and family members seek.

Conclusions: The results from this study identify specific strategies for health information technologies to better promote empathic communication when they communicate concerning health news. We distill the findings from our study into design hypotheses for ways technologies may be able to help people better cope with the possibility of receiving bad health news, including tailoring the delivery of information to the patients' individual preferences, supporting interfaces for sharing patients' context, mitigating emotional stress from self-monitoring data, and identifying clear, actionable steps patients can take next.

(JMIR Hum Factors 2019;6(2):e8885) doi:10.2196/humanfactors.8885

KEYWORDS

mobile health; eHealth; mHealth; patient-centered care; health communication; empathy

Introduction

Motivation

The proliferation of health technologies—such as fitness trackers, self-monitoring tools, and personal health records (PHRs)—enables people to be aware of their own health information more than ever before. The information patients may gather about their health from such technologies includes a casual notice of weight gain or loss, changes in cholesterol or blood glucose levels, signs of developmental delays, or an increased risk of a serious disease such as diabetes or Alzheimer's. Having access to personal health information via various technology channels can help people manage chronic conditions, encourage healthy habits, or bring awareness to problems they might not have previously recognized. Although people have frequent and fast access to their health data, the tools have the possibility of communicating bad health news without consideration for the patients' emotional condition to which a skilled clinician can be responsive. For example, with PHRs, patients can check their laboratory results on the Web without the presence of clinicians [1]. In the absence of the human element, such as the informational or emotional support that can take place during communicating health news in-person by a skilled physician, people could have difficulty assimilating information and making informed decisions about treatment options, lifestyle changes, and medications that could create undue emotional burden on patients. These situations could be avoided if health technologies are designed with *empathy*, which is known to positively influence patient health outcomes such as patient satisfaction and adherence to treatment [2].

In this paper, we argue that health information systems that potentially communicate bad health news need to deliver the news while considering the emotional needs for patients and that such needs have been largely unfulfilled in the design of current health information systems [3,4]. By investigating *how clinicians communicate bad news* about health, we can learn and apply strategies for designing health information technologies that are more empathic and can reduce the patients' emotional burden. In this paper, we first review established guidelines and protocols for communicating bad news that are designed to train clinicians to improve their communication skills. We then discuss the semi structured interviews we conducted with clinicians and patients to understand their respective experiences of delivering and receiving a diagnosis for severe or chronic conditions—such as cancer, Parkinson's disease, or diabetes. We identify and characterize the issues around health technologies that potentially cause patients anxiety, distress, and frustration and identify patients' and their caregivers' emotional and informational needs at the time of receiving "bad news." We discuss design hypotheses and example designs that leverage the strategies suggested by participants and guidelines from patient-clinician communication literature [5-10].

Strategies and Technologies for Communicating Health News

Clinical Guidelines for Communicating Bad Health News

Although communicating bad news is an important part of medical care, both clinicians and patients find it difficult. Clinicians have legal and ethical obligations to provide patients with as much information as they want [11] even if they suspect that it will have a negative impact on patients. A majority of patients desire to be told the truth about the diagnosis of a serious disease (eg, cancer) and even a grave prognosis [12]. As clinicians find it challenging to be honest with their patient and not destroy the patient's hope at the same time, many guidelines recommend how to communicate bad news.

Guidelines for communicating bad news are developed on the basis of reviews of other literature [7,13] and clinical opinions [14,15]. Although rare, a few studies account for patients' opinions [7,16]. Some guidelines are geared toward specific medical situations—such as communicating to cancer patients [14] or parents of a child with additional needs [17,18]. However, in general, communication skills are not disease-specific knowledge, and thus established guidelines can be applicable to a wide variety of situations where clinicians across specialties communicate with patients. Communication guidelines comprise ways to set context, listen to patients, acknowledge their emotions, and share medical information. It has been found to be useful in all medical interviews—especially in palliative care and psychotherapeutic dialogue—as the "breaking of bad news is universal to medicine" [8].

Communication guidelines and models assume that communication skills can be taught and acquired. Since the 1990s, North American medical schools began to teach communication skills. According to a 1999 survey in which 89 of the 144 medical schools participated, 85% reported that they teach communication skills [19]. Of the schools that used a structured model in teaching communication skills (32%), 2 models they commonly used were the SEGUE (short for *S* et the stage, *E* licit information, *G* ive information, *U* nderstand the patient's perspective, *E* nd the encounter) framework for Teaching and Assessing Communication Skills [20] and the Calgary-Cambridge observation guide [21]. As these models were general communication models, we looked for communication models specific to breaking bad news that are widely used in the medical community—the SPIKES model [14] and Consensus Guidelines [7]—and used them for framing our interview guides and analysis. The SPIKES model is useful for its simplicity, and the Consensus Guidelines are useful because of their comprehensiveness. The SPIKES 6-Step protocol emphasizes the sequence of communicative acts occurring alongside a process of emotional acknowledgment and repositioning. It comprises the following steps: (S)—Setting up the interview; (P)—assessing the patient's Perception; (I)—obtaining the patient's Invitation; (K)—giving Knowledge and information to the patient; (E)—addressing the patient's Emotions with empathic responses; and (S)—using Strategies and Summary [14]. Detailed strategies are provided within each step—strategies for "setting up," for example, include arranging

for privacy, involving significant others, and managing time constraints and interruptions; strategies for "obtaining the patient's invitation" refer to the process of determining how much information a patient wants to know and when they want to hear it. This guideline is based on the grounds that everyone has different information needs and that clinicians should ask questions (eg, "How would you like me to give you the information about the test results?") to gauge how much information a patient wants to know. The SPIKES protocol has been incorporated into a variety of training programs for clinicians and medical students across many disciplines. It has been evaluated by patients according to their rating of the procedure, perception, and satisfaction [22].

One caveat with the SPIKES protocol is that it is developed on the basis of communication techniques rather than empirical evidence. The consensus guidelines [7] on the other hand take a different approach of reflecting the clinicians' and patients' opinions during the process of developing the model. After a critical review of the medical literature on how to communicate bad news, the authors developed a draft of guidelines and then presented them to a consensus panel of medical professionals (n=28) and patients diagnosed with cancer (n=100) for their feedback. The consensus guidelines are a list of attributes rather than a sequence of communicative acts. They offer distinct guidelines such as being sensitive to patients' cultural, religious, or social background, employing a trained health interpreter if necessary, encouraging the patient to express his or her feelings and documenting what the patient has been told. Textbox 1 summarizes the consensus guidelines [7] for communicating bad news.

Textbox 1. Consensus guidelines.

Summary of recommendations for communicating bad news

- One person only should be responsible for breaking bad news
- The patient has a legal and moral right to information
- Primary responsibility is to the individual patient
- Give accurate and reliable information
- Ask people how much they want to know
- Prepare the patient for the possibility of bad news as early as possible
- Avoid giving the results of each test individually, if several tests are being performed
- Tell the patient his or her diagnosis as soon as it is certain
- Ensure privacy and make the patient feel comfortable
- Ideally, family and significant others should be present
- If possible, arrange for another health professional to be present
- Inform the patient's general practitioner and other medical advisers of the level of development of patient's understanding
- Use eye contact and body language to convey warmth, sympathy, encouragement, or reassurance to the patient
- Employ a trained health interpreter if language differences exist
- Be sensitive to the person's culture, race, religious beliefs, and social background
- Acknowledge your own shortcomings and emotional difficulties in breaking bad news

The review of the guidelines reveals a considerable overlap between SPIKES and the consensus guidelines such as ensuring privacy, assessing the patient's understanding of the situation, and providing an honest diagnosis using simple language. In a clinician's attempt to understand what it is like to be a patient, active listening and expression of feelings are the hallmarks of empathy during clinician-patient communication.

In our research, we used a review of the clinical guidelines to help frame our interview guides and as a starting point for our thematic analysis.

Patients' Preferences for Communicating Health News

Several studies have investigated patients' preferences for receiving health news, specifically in the context of receiving cancer diagnosis during the in-person communication [23-25]. For example, Parker et al conducted a survey to understand the characteristics of communication that different types of cancer patients would prefer such as what and how much information to receive, what setting and context they want to be in, and whether to receive emotional support during the communication [25]. Although these studies generate useful suggestions for improving in-person communication (eg, "Establishing a basis for breaking bad news" [23]), our goal is to identify insights for technology design. In this regard, Choudhry et al provide intriguing findings from their study on patients' preference for receiving skin biopsy results (which might contain a malignant diagnosis)—majority of patients (67.1%) preferred to receive the news via a telephone over other methods such as face-to-face communication (19.5%) or patient portal (5.1%) [26]. Top 2 contributing factors were (1) wanting to receive the results in the most rapid manner and (2) wanting to have an opportunity to ask questions when needed. In designing technologies for communicating bad news, we believe that these 2 aspects are important design considerations that need to be supported.

Self-Monitoring Tools for Health

Self-monitoring tools for health—such as blood glucose meters, electronic scales for body weight and body fat percentage, devices for sleep behavior patterns, and journaling tools for food—have proliferated in recent years. These self-monitoring tools often help people increase awareness of their behavior, identify patterns of behaviors, manage chronic conditions, or observe the effects of treatment. Self-monitoring tools could also improve the chances of early detection of a disease, which could also increase the chances of successfully treating it [27]. The real benefit of self-monitoring comes from using it on a regular basis long enough to identify trends. However, tracking data over time could cause anxiety when the data do not meet the observer's expectations, when the data show that the user is out of the normal range or if the user misinterprets or makes incorrect inferences from data [28]. Recent research has explored the phenomenon of people bringing self-monitoring data to their provider, but that presents a number of challenges [29,30] such as increased burden for both patients and providers, privacy concerns, and perceived disruption of a provider's primary care duties.

Personal Health Records and Electronic Medical Records

PHRs allow individuals to take an active role in managing their health and keeping their health information up-to-date [31]. Integrated PHRs—often referred to as patient portals or tethered PHRs—include a subset of health data from electronic medical records (EMRs) and provide more diverse features than an independent PHR. For example, they allow patients to access their laboratory test results, schedule appointments, or request prescription refills. Currently, the types of information that should be shared and how the information should be released have been the subject of heated debate [32]. Some clinicians are not enthusiastic about patients' direct access to their health information—such as laboratory test results and doctors' notes [33], despite their legal and ethical obligation to provide information if a patient asks for it. Clinicians worry as health information shared on the Web could potentially convey bad news to a patient, and thus patients run the risk of anxiety either with too little information (because of limitations of electronic media) or overwhelming information (in case of abnormal test results that may be difficult for a nonexpert to interpret). However, patients desire to have direct access to health information, including normal and abnormal test results, in less time than current norms [34]. Patient advocates argue that patients' direct access is a quick and efficient way of sharing information and might improve patient understanding and involvement in care [35,36]. Although we argue that health information systems should provide patients with direct and timely access to their own health information, this study offers design considerations for interfaces to minimize some of the negative consequences of such access on the patients' side.

Affective Computing

Affective computing approaches consider empathy as a physiological or behavioral measure and interpret those measures as emotions [37-39]. Studies in the affective computing literature often describe agent-based systems with animated humanoid software that emulates empathy through verbal and nonverbal modalities in various contexts. Agent-based systems are designed to alleviate a computer user's frustration [40], deliver discharge information in place of clinicians [41], or reduce stress levels of job interviewees [42]. Studies indicate that computers with such abilities can draw positive user reactions and increase people's desire to continue using the system. However, other studies show that agent-based systems are not yet sophisticated enough to replicate the subtlety and complexity of human empathy [43]. Boehner et al [44] assert that design should shift "from helping computers to better understand human emotion to helping people to understand and experience their own emotions." Although affective computing approaches concentrate on designing relational agents that emulate empathy, we aim to uncover opportunities for health technologies that support an empathic human-human relationship.

Emotional Support Through Health Technologies

We note that evaluation measures for health information systems are heavily weighted toward traditional usability (eg, screen layout) and efficiency (eg, learning ability, cost-effectiveness, task completion time, and error rate) aspects [45,46], and they often neglect how the system supports patients' emotional and mental states [47], though a recent study by Suh et al included emotional burden within their User Burden Scale for computing systems [48], and Kientz et al described considering emotional impact in the design of persuasive technologies [49]. For information and communication technologies studied in hospital settings, designers aim to improve clinicians' work efficiency or data entry [4,50], but they often neglect to support the emotional needs of patients. One exception is the study by Toscos et al, which highlights the importance of considering diverse emotional needs when designing health-monitoring technologies for teens with diabetes and their parents [51]. Technology has great potential to provide space for patients' emotional support. Researchers describe empathy as common in online patient support groups where patients seek both emotional and informational support [52,53]. Others reveal various characteristics of empathy presented in online discussion boards [54], or they have designed virtual agents to help convey empathy toward patients in care settings [41]. We can learn from these existing tools in the design of new health technologies.

Aim of the Paper

The primary goal of this research is to understand the design requirements for and investigate specific strategies for improving consumer-facing health technologies to communicate health news to patients in a way that is more empathetic and in line with best practices from clinical work in this space. The development of these requirements and strategies requires an empirical understanding of experiences of patients, clinicians, and patient family members.

Methods

Interviews With Clinicians and Patients

To understand the design space of using technology to communicate bad health news, it was critical for us to have firsthand dialogue with those who are involved in the process of delivering and receiving news about one's health. We thus conducted semistructured, open-ended interviews with clinicians, patients with chronic conditions, and patients' family members to better understand their experience and enable us to translate the findings from the medical guidelines into more practical considerations for our own work. Researchers from the medical field have conducted interview studies involving patients and patients' family members; however, these studies were aimed at developing guidelines for the clinicians [7,14], whereas our interview study is aimed at identifying opportunities for health information technology design. Moreover, clinicians' views (eg, feelings, thoughts, and behaviors clinicians have when delivering bad news) were studied mostly using structured surveys [14,55,56]. Therefore, it was important to include interviews with the 3 key stakeholders—patients, patients' family members, and clinicians. We chose to do a retrospective perception study rather than a study based on direct observation of clinician-patient communication as we considered asking patients or family members how they felt immediately after receiving bad news to be unethical and impractical. Although studies focusing on the communication of bad news are typically based on retrospective recall [57], we acknowledge that this approach has limitations—such as recall bias.

Recruitment

We recruited participants through word-of-mouth sampling and Craigslist postings in the United States. We interviewed a total of 23 participants—8 clinicians, 1 medical student, 1 social worker, 9 patients, and 4 patients' family members (see Table 1). Throughout the paper, we use the following naming scheme: "Cx" for clinicians, "Px" for patients, and "Fx" for family members. We offered a US $20 gift card to interviewees in appreciation for their participation. During screening, we sought clinicians or social workers who regularly conducted in-person medical diagnoses, prognoses, or consultations with patients. The social worker in this study had 12 years of experience in delivering the news of positive HIV tests to clients, and thus added a broader perspective than those trained as MDs (Doctor of Medicine) or nurses. We also sought patients who had been diagnosed with severe or chronic conditions. Although we did not formally define "severe or chronic conditions" in the recruitment posting, we listed cancer, Parkinson's disease, and diabetes as examples of these conditions, and we let patients self-identify what they considered as severe or chronic conditions. As there is limited literature reflecting the perspectives among clinicians, patients, and family members, we chose to include all 3 participant groups in this study. In addition, as empathic communication is universal across different conditions in health care, we expected that a diverse sample would give us insights into the variety of ways it manifests.

Table 1. Demographic details of participants.

ID[a]	Group	Age	Gender	Area of expertise or type of condition	Years, months, or weeks of experience
C1	Doctor	62	M[b]	Oncology	25 years
C2	Doctor	59	M	Women's health	25 years
C3	Med. student	32	M	Internal medicine	4 years
C4	Nurse	34	F[c]	Intensive care unit	3 years
C5	Doctor	45	M	Family medicine; psychiatry	19 years
C6	Doctor	39	M	Pediatric cardiology	11 years
C7	Doctor	__[d]	M	Internal medicine	14 years
C8	Nurse	—	F	Family health nurse	13 years
C9	Doctor	34	F	Internal medicine	10 years
C10	Social worker	45	F	Delivering HIV test results	12 years
P1	Patient	50	F	Parkinson's disease; breast cancer	17 years
P2	Patient	39	F	Diabetes; gastroparesis	27 years
P3	Patient	56	F	Follicular lymphoma—stage 4; diabetes	5 years
P4	Patient	57	F	Parkinson's disease; breast cancer; knee replacement	12 years
P5	Patient and Family	45	M	Himself—Thyroid cancer, Wife—ovarian cancer	Patient: 1 year; Caregiver: 2 years
P6	Patient	21	F	Heart disease	2 years
P7	Patient	34	F	Muscle disease (peripheral myopathy); Crohn's disease	1 year
P8	Patient	60	F	Uterine cancer; breast cancer	12 years (uterine cancer); 8 years (breast cancer)
P9	Patient	22	M	Bone cancer (Ewing's Sarcoma)—stage 4	1 year
F1	Family	45	F	Partner of P3	5 years
F2	Family	43	F	Sister was diagnosed with diabetes	2 weeks
F3	Family	43	F	Mother was diagnosed with liver cancer (stage 4) and passed away	1 month as a caregiver
F4	Family	35	F	Son was diagnosed with type 1 diabetes when he was 12 months old	3 years as a caregiver

[a]Naming scheme: "Cx" for clinicians, "Px" for patients, and "Fx" for family members.

[b]M: identifies male.

[c]F: identifies female.

[d]Age of provider was not given.

Interview Protocol

During the interview, clinician questions addressed the following: (1) perceptions of bad news, (2) diagnosis process, (3) strategies to deliver bad news, (4) common patient reactions and their coping strategies, and (5) perspectives on empathic care. We modified the interview questions for patient and family participants and asked the following: (1) the moment they heard the bad news and how it was communicated, (2) thoughts and reactions in receiving the news, (3) ways to manage and reduce distress, (4) the role of family members, and (5) memorable encounters with clinicians, either good or bad. All interviewees were encouraged to walk us through a specific case. Of the 23 interviews, 8 were conducted in person and the rest via phone. Interviews lasted from 30 min up to 2 hours.

Analysis

We audio recorded and transcribed all interviews to aid with analysis. We employed cross-case analysis of the transcripts using a thematic analysis approach [58]. During the interpretation phase, 2 researchers independently read through the transcripts and identified themes. The researchers then vetted, defined, and merged the themes into 1 code set. Using the preliminary code set, the 2 researchers independently coded the transcripts using Text Analysis Markup System [59]. Overall, 2 researchers exchanged the coded transcripts and reviewed the other's codes. The research team met regularly to discuss new themes and refine preexisting categories in the code set, thereby iterating on the codebook. The final, high-level categories of the analysis were characteristics of bad health news, strategies that clinicians use to express empathy

(understanding and communicating), patients' experiences and reactions in receiving bad news, patients' perspectives on poor communications of bad health news, and information and emotional support for patients and family members. We then used the analysis of the interviews combined with our review of clinical guidelines to develop our design guidelines for interactive technologies.

Results

Characteristics of Bad News

Bad news in the context of medical situations is defined as "any information which adversely and seriously affects an individual's view of his or her future" [60]. Bad news is in the "eye of the beholder," such that different people receive it differently depending on their life experience, personality, spiritual belief, philosophical standpoint, perceived social support, and emotional hardiness [57]. Clinician participants defined bad news as patients having a very serious illness, disease with poor prognosis [C6], or problem associated with the illness (eg , suddenly becoming blind from diabetes) [C5]. How people perceive bad news is context dependent. For example, bad news could be perceived as more tragic in young patients [C5], such an unexpected health condition affecting an infant, as opposed to the same condition affecting an older adult who already experienced related conditions. Moreover, not all bad news is perceived as tragic; if a disease is treatable or easy to manage, bad news could be heard as good news. P5 described as follows:

> *I had only thyroid cancer, not the lymphoma, which is very good news.* [P5]

Some participants [P4 and F4] even felt a *sense of relief* when they finally got a concrete diagnosis of a disease. On the other hand, a clinician being uncertain of what the patient has evokes anger and frustration on the patient's side. For example, P8 had a muscle disease, but her doctor did not know what type of muscle disease she had, even after many laboratory tests. This situation was frustrating for P8 as she did not know how to tell other people what medical condition she had or with which support group she could connect. However, a clinician participant had a different view. According to C9, not having a concrete diagnosis could turn out to be good news after all:

> *There are a lot of times when the diagnosis isn't sure, and that usually has a better prognosis. If I had a weird symptom, I'd prefer not knowing what it is, because chances are, it's not that bad in terms of statistics. It's counterintuitive, I agree. It's not the way we think, usually. But that's only because I know we've done the right tests...ruled out the bad things. Chances are, it's getting better.* [C9]

As such, how people perceive bad news is different for every person. Clinicians describe "bad news" in the objective sense on the basis of the severity and prognosis of the disease. On the other hand, patients and family members respond to bad news rather subjectively depending on many factors—such as past experience, expectation, personality, and religion.

Clinical Empathy and Empathic Communication

Definition of Empathy and Characteristics of Empathic Communication

The clinicians' empathic communication skill was particularly important in delivering bad news for both clinicians themselves (eg, a decreased risk of litigation) and patients (eg, lessening the distress). Clinicians in this study described empathy, in many ways, such as how C3 described it:

> *Understanding how you would feel if you were in the same situation as somebody that is going through an illness.* [C3]

C6 described it as:

> *Humanizing the diagnosis and the procedure* [C6]

C4 described it as:

> *Treating people like human beings rather than treating people like an illness.* [C4]

Finally, C3 described it as the following:

> *A clinicians' empathic communication skill is "more like an art than a science".* [C3]

A clinician's empathic communication skill requires the ability to create a connection with people that is beyond just clinical information. When we asked clinician participants if being empathic to patients can be learned, many agreed that empathic communication is indeed a learnable skill.

Experienced clinicians are well aware of the intrinsic value of empathic communication—the recursive process of *understanding* and *communicating* with patients—which is different from the step-by-step process that the SPIKES protocol suggests [14]. Empathy is hardly ever communicated without the clinician's understanding and acknowledgment of the patient's context. For example, the clinician might need to know the patient's feelings, level of understanding of the disease and options, work situation, and home life. Furthermore, the clinician's understanding of a patient's situation and emotional state means little unless the clinician is able to skillfully communicate that understanding. Understanding and communicating happen simultaneously as clinicians consciously and continuously reassess the patient's situation. Confirming the guidelines, clinician participants said they modify their method of delivering unexpected news on the basis of the patient's feedback and life story. However, C8 stated that modifying the method of delivery is often hard to achieve in the intensive care unit where patients rely on a ventilator and other supporting devices and often cannot communicate directly with clinicians.

Strategies to Understand Patients' Context

We identified that empathic clinicians make an effort to understand patients' context *before* and *during* the patients' visit.

Preparing for the Patients' Visit

Before meeting with patients, clinician participants reported a need to remind themselves of the patient's situation by checking the patient's chart, reviewing information, and looking for

certain characteristics (eg, the disease, laboratory results, records of previous procedures, or other key events). Patients' occupation or cultural background was additional information that helped clinician participants adjust their way of speaking to accommodate patients' medical understanding.

Anticipating Patients' Feelings Through Careful Observation

In preparing to deliver unexpected and life-changing news to patients, clinician participants reported not only anticipating the patients' level of medical understanding but also acknowledging patients' feelings. The 5 stages of grief model (denial, anger, bargaining, depression, and acceptance) by Kubler-Ross [61] was often referenced during the interviews with clinician participants when they explained the importance of knowing where patients are in their feelings. Knowing where patients are in this model by looking into patients' eyes helped clinician participants assess patients' feelings and gauge what information to reveal when. C5 and C6 described as follows:

> *You have to watch them, and watch their faces. You have to try at least to read and get a feel for where they are with the conversation, is the first step.* [C5]

Another clinician said the following:

> *So I try to tell them as much as I can...but I gauge it on the family and the parents, and I try to watch them and look at how much I'm giving them and how they're reacting because it can be...it's very overwhelming.* [C6]

The clinicians we interviewed stated that knowing where patients are at emotionally helps them work around the state of shock and anxiety that often prevents patients from fully absorbing critical information. When clinicians perceive that patients are emotionally charged, clinicians might step back and wait for a better time to reveal certain information, invite patients to call with questions, or suggest that peers and family be present to help ask questions or make sense of the information. In this sense, the clinicians' ability to empathize with patients is what helps the clinicians aid patients in assimilating troubling information.

After achieving an *understanding* of patients' context through pre visit preparation and anticipation of patient's emotional state through careful observation during a consultation, clinicians should be able to skillfully communicate that understanding. In what follows, we describe the strategies clinicians use to *communicate* with patients, which are the other important part of empathic dialogue.

Strategies to Communicate With Patients

Communicating empathy refers to clinicians' acknowledgment of patients' feelings. Clinician participants described several communication techniques they use to convey empathy while presenting information directly and simply, which aligned well with the clinical guidelines we analyzed.

Building a Partnership of Trust With the Patient

Clinician participants reported they commonly use their opening statement to reinforce a partnership. Clinicians want patients to trust in the quality of their care. Trust between patient and clinician alleviates patient fear, which could smooth the decision-making process that must occur around every new piece of clinical health information. To build a partnership of trust with patients, several clinician participants mentioned using language that reinforces an "us" relationship rather than a clinician versus patient hierarchy. The following examples show how clinicians reinforce a partnership, as stated by C1:

> *All of us are advocates for the baby and you.* [C1]

And as reported by C2:

> *I'm glad you came. Let's look at that report. Let's look at it together.* [C2]

And finally, by C5:

> *I'm gonna have to tell you something that's difficult and I'll give you all the details so that you understand it. I want you to know that we'll work with you to make sure you really fully understand it.* [C5]

Acknowledging Physical and Emotional Discomfort

Another way to communicate empathy is to address the patients' feelings directly. Clinician participants reported using comments such as the following provided by C2:

> *It must be hard to encounter something that may seem so serious...I am sorry you are in pain. I hope we can work to make you to feel more comfortable.* [C2]

And the following by C5:

> *I could imagine how frightening this is to you.* [C5]

However, 1 clinician participant expressed the difficulty of having to maintain a certain distance from patients but wanting to empathize with their feelings at the same time:

> *It can be very tough if you become emotionally involved with the patient. For me personally, I try not to get completely tied in with them, but at the same time, I don't want to not be saddened by telling a parent that their child is going to or has died. If I ever get to a point when I have a conversation with a family delivering them news of a prognosis and it doesn't affect me, that would worry me that I'm too disconnected.* [C6]

Others observed how even in situations where a clinician could not save a patient's life, the clinician's empathic acknowledgment of the difficult situation made family members feel that they were being treated as human beings.

> *As I remember, hearing her deliver those news in such a loving, caring, compassionate way..."I care for you, I'm saying something that is very hard, I will be with you, there's nothing we can do really to avoid the ultimate result, but we will work together to make it the best for you that we can." And she was true to her words. She was there all the time. And we could thank her for having been there with us for 2 years.* [P5]

Although P5's wife passed away, P5 was grateful to the wife's clinician for the empathic approach to providing patient care. As such, a clinician's acknowledgment of patients' and family

members' emotional feelings helps them deal with bad news and go through tough times.

Setting up a Scene Where Patients Can Process Information

Creating a space for empathic dialogue between clinicians and patients requires that patients be in a comfortable and private environment where they can process the information being conveyed. A clinician participant [C9] described how she prepares to communicate bad news to an inpatient. After she makes sure that the patient is in a private environment, she does the following:

> *There's usually a short social phase, where you talk to the person about how they stay at the hospital...you find something to make everyone feels at ease, you make sure whether they are sitting comfortably...you sort of unconsciously check that there's tissue somewhere in the room if it's really bad that you are gonna be announcing. Um...there's usually tissue in your pocket or something...you know, that might be an issue...having to get up to go find tissue is not as nice afterwards. So as much as you can plan before, but that's just a small thing that you just learn with time. That's not in the textbook.* [C9]

According to our participants, the actual diagnosis is the most important piece of information for patients. The same information could be delivered in various ways—from people in different positions using different means of communication, and those ways affect the conveyed empathy. One patient participant received an unexpected phone call from a nurse saying the following:

> *Hi, we just wanted to let you know that the biopsy came back and it is a cancer.* [P4]

Others were informed by an experienced clinician who carefully revealed the diagnosis along with descriptions of the condition. The clinician then opened a dialogue wherein the patient and clinician could discuss treatment options, prognoses with and without treatment as well as what the patient could expect to go through with surgical procedures, side effects, expectations for healing, and lifestyle changes. The clinician's goal in creating a time and space for empathic dialogue is to ensure that a patient fully understands his or her condition to make informed decisions without becoming overwhelmed in the clinical details.

Building Resilience and Giving Hope

The experienced clinicians described the importance of developing the patient's emotional strength as that is what makes patients endure painful or chronic conditions; a clinician stated the following:

> *...the will to fight* [C1]

Even though it is discomforting for clinicians to tell patients the following:

> *This is what you will die from* [P5]

All clinicians we interviewed stated the importance of being honest, clear, and straightforward when delivering diagnoses. What is more important yet difficult is to obtain the balance between being honest about a poor prognosis and giving hope at the same time. Giving hope is different from giving false hope, which several participants also referred to as "sugarcoating." Sugarcoating is telling patients glossy stories and assuring them that everything will be fine when in fact the patient is in failing health. All clinician participants asserted that sugarcoating is harmful for patients, and it only protects clinicians who want to avoid dealing with the patient's emotions. However, giving hope helps in a situation when patients have to develop both the physical and emotional strength and resilience to endure a difficult situation. The gynecologist we interviewed told a story in which he encouraged a cancer survivor to consider undergoing a high-risk surgical procedure that would dramatically alter her physically but would also extend her life. He stated the following:

> *I looked at this woman—tremendous will, tremendous spirit—I brought her back and said, "There's something that can be done. It's a very radical surgery, and not many survive it. But those who do, they do well. So I need you to consider this. It means having an operation to remove ovary, bladder, vagina...all of them will be cleaned out...you will be sick, you will be in hospital for many days...but you may live. I think that you are tough and you can make yourself come through this. Are you up to this challenge? I think you can do it. I think you have it in you." In turn, this patient needed one more chance of hope. So she got operated, and she survived.* [C2]

The quote above illustrates not only the clinician's confidence in the patient's capacity to endure a radical procedure on the basis of his previous knowledge of the patient's life story and medical history but also the level of trust in the clinician-patient relationship that allowed him to speak with honesty and candor about the surgical outcome. Clinicians' knowledge about their patient's life story and their ability to communicate with such candor and trust helps patients build resilience and hope, which are indicative of empathic dialogue.

Matching the Level of Information to the Patients' Level of Understanding

The experienced clinicians we interviewed present complex information in plain language to do the following:

> *...make the person in charge of their situation.* [C2]

In addition, to give patients the following:

> *...good information so they can make good judgments about their lives.* [C1]

When explaining data, clinicians break difficult concepts into down-to-earth terms and use visual aids such as drawings, graphs, pictures, and x-rays. Another strategy the clinicians used was to tailor their language to the patient's life experience. For example, 1 clinician described speaking in probabilistic terms with patients who, as engineers, appreciated the mathematical explanation.

Communicating Face-to-Face, if Possible, and Using Nonverbal Means

All clinician participants explained the importance of face-to-face communication and being mindful of nonverbal

communication during consultation. They preferred face-to-face communication with patients in a quiet, private space where they could maintain eye contact and, if needed, sit beside patients to look at data together. Some clinicians said they do not allow sensitive information to be delivered over the phone or by staff members who do not know how to communicate empathically. C10, who worked at an HIV clinic, also mentioned that it was the clinic's policy to never give news over the phone regardless of the test outcome. Observing the patient's (or client's) body language and facial expression allows clinicians to tailor the way they give a message to individual patients.

However, face-to-face communication is not always possible if a time-critical test result comes back outside office hours or between the scheduled appointments. In addition, time constraints often limit the face time during appointments. Indeed, a few of our patient participants received their diagnosis over the phone, and some of them were grateful for their clinician's attempt to reach them as soon as possible when the information was urgent. A patient stated the following:

> *What I still appreciate about my doctor at that point is that she called me. She actually called me while I was at work. And she called and asked if I was alone. "(Name of P8), I have some news for you that the test showed that you have a uterine cancer." I appreciated her honesty, and that she called me. (...) She asked me if I was alone, and there was a part or piece in that she knew me well because I would not have wanted to get that information while other people were in the office and I wanted to focus on talking to her on the phone.* [P8]

In addition to this list of strategies we discussed, clinician participants also emphasized the importance of active listening, being responsive, and spending enough time with the patient. In practice, not all clinicians can employ these strategies when they communicate with patients because of time and resource constraints, which is where empathically designed technology might be able to help fill the gap. We next turn to patients' perspectives on what helps and does not help when they receive bad news.

Patients' Experience of Receiving Bad News

Patients' Reactions to Bad News

When people receive a diagnosis of a severe or chronic disease, either of their own, or of their family member's, their life changes in many ways. A patient might move to a bigger city for better care, whereas a family member might move closer to support the patient. We begin with describing a scenario of a patient who is about to learn her diagnosis. We reconstructed the scenario on the basis of P4's experience:

> *A doctor walks into a room, and he is about to tell a working mother of 2 that she has Parkinson's disease. The patient has been having trouble with small motor operations, such as unlocking a door with a key. She underwent MRI and CAT scans during a previous visit. She is waiting for the result, not knowing what the radiologist was looking for. She has been enduring a low-grade fear: fear of telling her coworkers and daughters, fear of losing her job because she is a construction inspector and her job requires driving, and fear that if she loses her job, she will also lose her health insurance coverage.* [Reconstructed on the basis of P4's experience]

As portrayed in the above scenario, patient participants expressed various kinds of fears that they experienced while waiting for a concrete diagnosis of a serious disease, including fear of losing their job, losing health insurance, having to rely on others, taking regular shots, and having to use a cane, pacemaker, or feeding tube early in life. Some patient participants also experienced fear of pain, death, progression of illness, and of situations such as being chased and not being able to escape because of their condition.

Although some patients described feeling shocked at receiving an unexpected diagnosis, others, because of their perspectives from previous challenging life experiences, did the following:

> *...took it all in a stride* [P3]

Moreover, patients who visited multiple clinicians expressed *relief* at finally receiving a diagnosis that was true to their symptoms and in knowing how to manage an illness or knowing the next steps to take. For example, it took 2 years for P1 to get a concrete diagnosis of Parkinson's disease. When she finally heard that she had Parkinson's disease (after seeing 10 clinicians), her first reaction was a great sense of relief.

Patients also expressed feeling suddenly different. Accompanying reactions include being angry with their bodies for not working and hiding their condition and emotions from coworkers, family, and friends. P2 described the following:

> *I try not to show that I'm in pain, and I try not to show that I'm not feeling good because it's just...I think it makes people feel bad to be around somebody that's just not feeling good.* [P2]

In the United States, finances and the cost of care are key factors in selecting a course of treatment, especially when care is costly (eg, intensive care unit) or when procedures are not covered by insurance. Patients and family members are cautious about revealing information about health conditions in the workplace, because if they lose their jobs, they may also lose subsidized health insurance. Both patients and family members are sensitive about with whom they share the bad news. F1, a partner of P3, was an executive director of an organization. She explained why she did not want to tell her colleagues about the partner's health situation after receiving bad news. She described it as follows:

> *...because of my role as executive director, every time that [Name of P3] was going through chemo, you know, I didn't want to tell my board of directors because I thought that they would think that that would impact my performance, and it was just something that I did not want to share....I didn't want to be seen as an absent executive director.* [F1]

Some patients and family members face workplace discrimination because of frequent absences and perceptions of lagging performance. Some patients can no longer work and must go on disability leave, which means adopting a new role

that is different from being an employee. If patients have a severe condition, they might have to rely on others to help with shopping or driving. They might lose the freedom to walk around by themselves. When patients cannot care for themselves, they stay in a hospital. F1 eventually shared the bad news with her colleagues, and she later even called all of P3's friends to let them know of P3's cancer and diabetes and asked for their support. However, the initial fear and emotional fall-out that patients and family members experience at the very beginning stage of care prevent them from actively asking for and receiving the support in a timely manner.

Patients' Perspectives on Poor Communication of Bad News

Patients and family participants had varying degrees of experience—in good ways and bad—in receiving bad news from clinicians. The bad experiences, in particular, were so hurtful and thus memorable that patients were able to articulate how they had felt when receiving bad news, although many years had passed since then.

Patient participants were irritated when the clinician was insensitive to their experiences and treated them like "just a number." Patient participants eventually became angry when the clinician did not listen or asserted his/her opinion over the patients' experience. Patient participants also expressed frustration with clinicians when the clinician did not offer sufficient opportunity to ask questions, did not answer questions, or did not adequately explain procedures, as P3 described the following:

> *He [doctor] definitely didn't make some things very clear, like you know, I was kind of scared to ask him why...why aren't you giving me these tests, why wouldn't you give me these tests if I'd had the money.* [P3]

Patient participants reported that "bad doctors" are "cold," "pompous," and "callous" clinicians who are perceived to avoid dealing with patients or put the responsibility for communicating with patients on somebody else, disregard patients by treating them as subordinates, and prioritize clinicians' own interests over patients' needs. Patient participants were especially frustrated when clinicians did not spend enough time with them. P2 and P5 shared one of their experiences of receiving bad news from "bad doctors." P2 said the following:

> *"I don't have time right now," she [the doctor] said, "talk to one of the nurses. They can answer your questions. I'm too busy..."* [P2]

P5 stated the following:

> *The very first interaction learning about it [cancer] was this very ridiculous setting which he [the doctor] was standing by the door, just to ready to leave, and saying "Oh...by the way I forgot to say, you have a cancer." I could kill him.* [P5]

Poor communication of bad news left patients with more questions than answers and caused patients to withdraw and assume the issues were internal and somehow their fault. Some patient participants experienced depressive symptoms such as denial, withdrawal, and suicidal ideation. All patient participants that we interviewed, at some point in their lives, encountered clinicians who did not have a good bedside manner or empathic communication skills. When patients felt their clinician is not on their side, they sought second opinions or eventually switched clinicians. Patients also turned to other sources of comfort and built lifelong relationships with those having similar conditions, which is what we will discuss in the following section.

Patients and Family Members Seeking Emotional and Informational Support

After receiving a diagnosis, patients and family members sought emotional and informational support to cope with their medical condition and distress. At the time of diagnosis, it was hard for the patients and family members to know what questions to ask clinicians. In addition, clinicians often did not provide enough information, or even if they did, patients and family members are overwhelmed by the amount and content of the information, and they have a hard time assimilating it. However, as time went on, patients and family members became researchers and sought information from other sources—such as books and the internet—besides their clinicians. A patient stated the following:

> *And so I study a lot. I go to the library, and just look at research magazines and books that are anything related to diabetes and complications and anything like that. I get most of [it] from the internet.* [P2]

Another said the following:

> *So I'm thinking, how can I assist? I went out and, well, they're delivering, Amazon.com, I ordered Diabetes for Dummies, and I did go with her to the meeting with the nurse and dietician...* [F1]

Local support groups were also a great source of information. Patients initially learned about support groups from clinicians, hospital waiting room materials, and associations' websites. In the support group meetings, people shared an enormous amount of information that could only be learned from experienced patients who "have been diagnosed with this." Patients talked about clinicians, procedures, drugs, and complications and obtaining information that clinicians do not give or cannot answer. Patients meet other patients from similar age groups, share their experiences, and make lifelong friendships. Patient participants described organizing special events such as children's workshops, fundraising events, summer camps, galas, and dancing.

However, not all patient participants saw support groups so positively. A patient stated the following:

> *I went to the support group, but I was in denial. I mean I wasn't accepting the fact that I was having this [Parkinson's Disease], and I wasn't telling anybody, and I went to a support group meeting. There were way too many people, way too overwhelming. And I didn't like seeing the various stages of people with Parkinson's. So I didn't go back until October of last year...* [P4]

As P4 mentioned, being able to project how his/her health will deteriorate by observing other member's conditions or being

notified that a group member had deceased could make patients and family members feel depressed, uncomfortable, and prevent them from actively participating in the local patient support group.

Online health communities such as online discussion forums, live chat rooms, mailing lists, and newsgroups were also popular sources of providing emotional support and health information, which confirms existing literature [62]. However, these online health communities and online support groups imposed similar problems to local support groups in that being notified of others' bad health conditions and their dramatic reactions could make the patient and the family member feel uncomfortable. F4 was a mother of a 4-year-old child with pediatric type 1 diabetes. She became an expert caregiver of her son, but she felt that online patient forums were not helpful for her anymore; she stated the following:

> *...it is not as good for me [to go to an online patient forum] because pretty much, all of those parents who just found out...they are still kind of shell-shocked...So it's not so much as a support group. Nobody slept, everybody is shell-shocked, and everybody is freaked out...it's kind of depressing.* [F4]

Information does not always equal comfort. If a patient's diagnosis is a rare or specific one or has a grim prognosis, information from the internet and online support groups that is not specific to the patient's situation might not be helpful, and it could even be sometimes harmful. P9 was diagnosed with Stage 4 Ewing's Sarcoma, which is a rare type of bone cancer with a very poor prognosis. P9 stated the following:

> *My doctors said, "Don't look it up. Don't go on Internet...because it is so specific to each person. Just ask us questions directly." And they were really good at providing me with answers. And when someone did [looked up on the internet], I took that information with a grain of salt, and said, "It's probably not specific to me." Because my cancer was stage 4, so it wasn't good from the outlook from the beginning. So the Internet was not helpful.* [P9]

Regardless of these drawbacks, online health communities and online support groups could be a critical place for patients and family members to share personal experiences and actionable advice to cope with day-to-day health issues [63]. However, our findings about the depressing or improper use cases of patient group websites call for careful design of these sites as the information offered by other patients can only be helpful if it is accurate and tightly relevant to the inquirer's situation.

Discussion

Principal Findings

Clinical guidelines for communicating health results exist to help clinicians identify strategies to help communicate bad news to patients in a way that puts patients' emotional needs first. Clinician participants in this study tried to follow these guidelines, and when they do, they are well received by patients and their family members. Thus, there is an opportunity to apply these strategies to the design of consumer health technologies. Below, we list several design hypotheses, as recommended by Hekler et al [64], for ideas for implementing better empathic communication within technology systems that potentially communicate bad news. We call them "hypotheses" instead of recommendations or implications as they require additional testing before they can be generalizable knowledge [64].

Design Hypotheses for Consumer Health Technologies That Communicate Concerning News

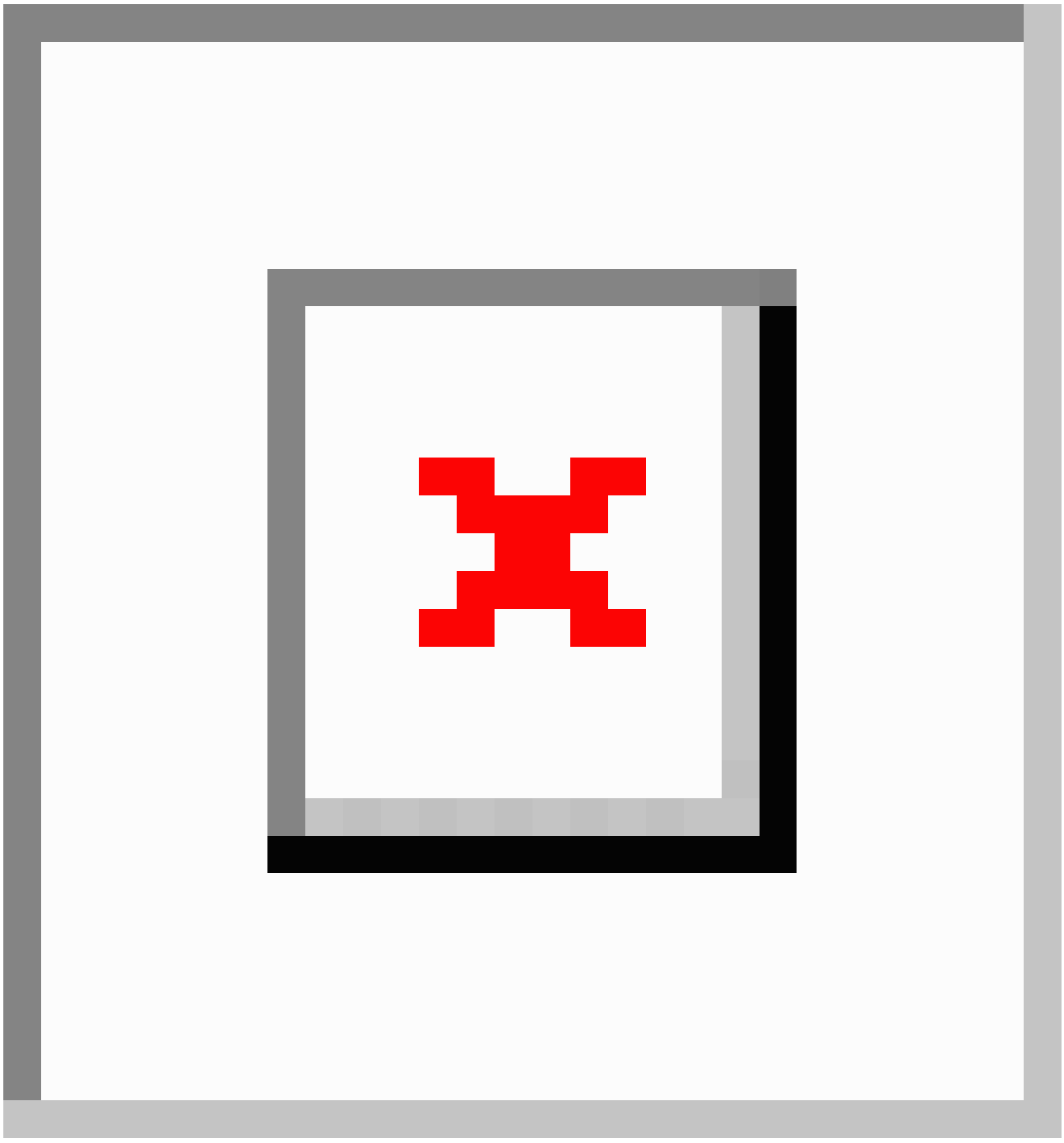
We acknowledge that not every clinical guideline can be applied in the design of health information technologies, nor do we believe that human practices can fully be facilitated by technology, but we believe technologies that may do this could be better designed. In this section, we provide a series of design hypotheses for how technologies could be designed to convey bad news and discuss how these specific design ideas can be applied to the design using health technology examples.

Design Hypothesis 1: Tailor the Delivery of Information to the Patients' Individual Preferences

Patients have different information needs and personal preferences (eg, how they want to be contacted by a clinician, whether they prefer participating in online/offline support group), and clinicians can ask the patient how they would like to receive information at the time of ordering the test (eg, face-to-face, via a PHR) and when they would like to receive it (eg, as soon as it's available, after the doctor has time to review it, etc). This aligns with the SPIKES guideline of obtaining the patients' invitation [14].

In terms of delivering laboratory and diagnostic test results through a PHR system, we believe patients should have instantaneous access to their results, without delay, if they choose. As our patient participants stated, having a concrete diagnosis brings patients a sense of relief even though the diagnosis may be a serious disease such as Parkinson's disease or pediatric diabetes. However, an information buffer could be placed in the system, which gives people the option to wait until a medical professional can help them accurately interpret the results with an explanation of terms (eg, the meaning of the medical terms, screening, sensitivity, and specificity) in the context of their specific health situation. For those who want to receive information verbally from a clinician, the system could send a note to the patient when the results are available and have the patient schedule a phone call or a visit. It should not be the intention to hide the information but to suggest a compassionate way of delivering a piece of potentially concerning health news by providing it at the right time. Moreover, patients need options to decide how and when they want to receive the news. A system could also provide a secure means (eg, email, voice mail) for patients to contact the clinician if they have any questions or concerns about the results and inform when the clinician will reach out to the patient. In addition, technologies could provide additional information from a trusted source where patients can begin to conduct their own research.

Figure 1. 23andMe, a popular genetic testing site, adheres to several guidelines for empathic communication of potential bad news. Upper panel: the site confirms with users that they are ready to see their health results. Lower panel: 23andMe gives options for speaking with medical professionals for further information.



Health Technology Example

23andMe [65], a service for genetic testing, has on their website a method for delivering sensitive genetic information about increased risk of Alzheimer's and Parkinson's disease that aligns with this recommendation (see Figure 1, upper panel) by asking users to confirm if they would like to receive their results or not, and it explains the risks before showing the results. They also offer information on how to talk with a genetic counselor to get more information on interpreting the results (see Figure 1, lower panel).

Design Hypothesis 2: Support Interfaces for Tailoring Toward Patients' Context

In this study, clinicians' understanding of context such as patients' feelings was an important part of empathic dialogue. Health information technologies could be designed as learning and prompting tools for clinicians to better understand patients. It was emphasized in many guidelines that it is essential for clinicians to gauge a patient's level of understanding and emotional state during the consultation before communicating bad news. The clinicians we interviewed mentioned they were already taking notes about patients' backgrounds and unique characteristics during medical consultations and read these notes

right before the next visit. In addition, patient participants appreciated clinicians who took the time to listen to their stories and family background, which often is not necessarily reflected in the medical chart. Therefore, it could be possible to have patients add their own notes about their emotion and their background to a specific section in their medical records through a PHR. Patients could complete an electronic form where they can detail their background (eg, family history, emotional state, and preference of receiving news) in advance of the visit or while they are waiting. Future designs could tailor this over time as a patient adapts and changes preferences. This would also allow clinicians to be mindful of the patient's emotional state or whether to invite close family members to the consultation, and it would provide an opportunity for more automated generation of tailoring news to participants' preferences that happen remotely.

Health Technology Example

The field of health communication has been successful in computer-based tailoring of messages in domains such as smoking cessation [66,67], weight loss [68], and mammography screening [24] on the basis of aspects such as cultural background, gender, stage of change, marital status, whether they have children, and their social support [24,66]. As a specific example, Stretcher et al [66] found that a simple smoking cessation website tailored on the basis of baseline questionnaires participants completed at the beginning of the study was more successful than generic messages. Similarly, tailoring messaging in PHRs on the basis of user preferences and context could be used to deliver potentially bad health information in a more empathic manner by meeting patients where they are and only sharing news they are ready to hear in a manner with which they are comfortable.

Design Hypothesis 3: Mitigate Emotional Stress From Self-Monitoring Data

Several empathic clinicians in our study attempted to build a partnership of trust with patients and acknowledge patients' physical and emotional discomfort. However, when people receive personal health information from commercial self-monitoring tools, they do not have a counterpart of a care provider who can provide emotional and informational support. While using self-monitoring tools, people might feel distressed when they find they have not met their weekly goals or when they feel what they have been experiencing is abnormal. Take an example of a patient who is experiencing severe pain after surgery and is monitoring his pain level. Feedback from a pain tracking system could convey information about what is normal in plain language (eg, "80% of people experience severe pain after this surgery") with an aim to lower the patient's distress. Interfaces could also use language that reinforces an "us" relationship similar to what our clinician participants stated. For example, when a glucometer presents a higher than normal blood glucose reading, the interface could say, "A single high blood sugar reading usually isn't a cause for alarm, but let's check a few things together," and guide the patient through possible reasons—medication, food, and exercise.

Health Technology Example

On the developmental screening results page for Baby Steps [69,70], we use language that acknowledges that it is normal to feel anxious about how your child is doing developmentally and provide some sense of what is normal, which might cause potential for worry but is not actually worrisome (eg, variation across categories, small plateaus, not answering "yes" to all screening questions). We also use "we" language to emphasize a partnership in tracking children's progress and working together to accomplish the task of monitoring children's development. For example, language describing how to interpret the visualization of the results states, "Rohan could use some encouragement in this area. Let's find some developmental activities to try with him." We also tested early screen mockups of different visualizations of the results for developmental screening with parents in a Web-based survey. The resulting visualization that received a high level of understanding of the results and also reduced anxiety was a more abstract visual metaphor to communicate the child's developmental progress where different sizes of trees represent the child's growth (see Figure 2). This visualization used the metaphor that children grow at different rates, and a lower score on a developmental screen may just mean that their child has not yet had the opportunity to grow in a given area. Currently, as there is no evidence on the fact that hitting milestones earlier has an impact later in life [71,72], we chose to only communicate results if a screen indicated children were at risk of developmental delay and needed further evaluation or needed to be encouraged with developmental activities rather than showing exact percentiles.

Design Hypothesis 4: Help Identify Clear, Actionable Steps Patients Can Take Next

Some patients in this study reported feeling helpless when they received bad health news that was communicated poorly and that they expressed a desire for things they could do to feel less helpless. Moreover, 1 way to accomplish this would be to help patients by giving them clear, actionable steps they can take after receiving a diagnosis. This could be as simple as giving them trusted information they can read more about, suggestions for contacting a close family member or counselor and instructions for what to say to get support, or actions they can do to start treatment, such as scheduling an appointment with a clinician.

Health Technology Examples

Overall, 2 of the previous technologies we described have good examples of this design recommendation in practice. For 23andMe [65], patients are given the option to talk with a genetic counselor directly through the site on the basis of the results of a genetic screen (Figure 1, lower panel). With Baby Steps (Figure 2, bottom), we couple results from a developmental screen with information for the parents immediately on the screen where they see the result. If the result that the child is close to the cut off for having a developmental delay, Baby Steps links parents to a list of activities they can do with their child that encourage development, which they can check off as they complete them. If the result is that they need an evaluation beyond self-monitoring, parents are linked to free services they can contact, which will help them to conduct a

more formal evaluation, and they are given the number to a toll-free parent help hotline they can use to talk to someone immediately.

Figure 2. Interface for conveying the results from a developmental screen in the Baby Steps Web portal. The different sized trees represent where a child is at developmentally for a given category. Immediately below the trees is an interpretation that uses team-based language, acknowledges the potential for anxiety, and indicates that variation is normal development.

Conclusions

The objective of our research was to uncover insights for the design of health technologies that potentially convey concerning news. We accomplished this goal by (1) examining established guidelines for clinicians on communicating bad news related to health, (2) conducting interviews with patients, patients' family members, and clinicians on their experience of delivering and receiving a diagnosis of a serious disease, and (3) rethinking the design of health information technologies—EMRs, PHRs, and self-monitoring tools—to support clinician-patient empathic dialogue and reduce the discomfort of patients when they receive bad news. We have addressed how the human element is conveyed during medical practice, especially when communicating diagnoses of severe or chronic diseases. We also identified how clinicians develop their own strategies to understand patients and communicate with them, and we

investigated patients' internal turmoil and emotional distress when receiving bad news and emotional and informational support that patients and family members seek elsewhere. We tied our findings to 4 design hypotheses for health technologies aimed to facilitate better self-managed care and promote the expression of empathy in the clinical setting, and we demonstrated their application in different health technology designs. We believe that future work might be to explore these design hypotheses and validate both positive and negative technology examples empirically with potential users as well as explore how strategies for empathic communication might evolve over time.

Empathic communication should be considered a core value in the design of health technologies [73], and a more empathic approach to design is needed [74]. Patients' needs and their situations are different and a "one-size-fits-all approach" does not work. However, health information technology has a great potential to support and reinforce the empathic relationship of a clinician and patient. Our approach of investigating the best-case practices of empathic communication is the first step to bringing "empathy" into the designs of empathic health information technologies.

Acknowledgments

The authors would like to extend their heartfelt gratitude to the participants who so openly shared their difficult experiences. This research has been reviewed and approved as exempt by the University of Washington's Institutional Review Board, protocol #35607. This study was funded by the National Science Foundation (#1344613, #0952623) and the Intel Science and Technology Center for Pervasive Computing.

Conflicts of Interest

JAK's spouse is the cofounder of Senosis Health, a startup company in the area of health technologies for diagnosis, monitoring, and treatment, which was acquired by Google in 2017. The remaining authors declare no conflicts of interest.

References

1. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
2. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. Eval Health Prof 2004 Sep;27(3):237-251. [doi: 10.1177/0163278704267037] [Medline: 15312283]
3. Siegler EL, Adelman R. Copy and paste: a remediable hazard of electronic health records. Am J Med 2009 Jun;122(6):495-496. [doi: 10.1016/j.amjmed.2009.02.010] [Medline: 19486708]
4. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. Br Med J 2004 May 15;328(7449):1184-1187 [FREE Full text] [doi: 10.1136/bmj.328.7449.1184] [Medline: 15142929]
5. Barton E, Aldridge M, Trimble T, Vidovic J. Structure and variation in end-of-life discussions in the surgical intensive care unit. Commun Med 2005;2(1):3-20. [doi: 10.1515/come.2005.2.1.3] [Medline: 16808704]
6. Brock CD, Salinsky JV. Empathy: an essential skill for understanding the physician-patient relationship in clinical practice. Fam Med 1993 Apr;25(4):245-248. [Medline: 8319851]
7. Girgis A, Sanson-Fisher RW. Breaking bad news: consensus guidelines for medical practitioners. J Clin Oncol 1995 Sep;13(9):2449-2456. [doi: 10.1200/JCO.1995.13.9.2449] [Medline: 7666105]
8. McFarlane J, Riggins J, Smith TJ. SPIKE$: a six-step protocol for delivering bad news about the cost of medical care. J Clin Oncol 2008 Sep 1;26(25):4200-4204. [doi: 10.1200/JCO.2007.15.6208] [Medline: 18757335]
9. Rao JK, Anderson LA, Inui TS, Frankel RM. Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. Med Care 2007 Apr;45(4):340-349. [doi: 10.1097/01.mlr.0000254516.04961.d5] [Medline: 17496718]
10. Reynolds WJ, Scott B. Empathy: a crucial component of the helping relationship. J Psychiatr Ment Health Nurs 1999 Oct;6(5):363-370. [doi: 10.1046/j.1365-2850.1999.00228.x] [Medline: 10827644]
11. Goldberg RJ. Disclosure of information to adult cancer patients: issues and update. J Clin Oncol 1984 Aug;2(8):948-955. [doi: 10.1200/JCO.1984.2.8.948] [Medline: 6747672]
12. Morris B, Abram C. Georgetown University Library. 1982. Making healthcare decisions: the ethical and legal implications of informed consent in the practitioner-patient relationship URL: https://repository.library.georgetown.edu/bitstream/handle/10822/559354/making_health_care_decisions.pdf [accessed 2019-03-06] [WebCite Cache ID 76fgYiimL]
13. Rabow MW, McPhee SJ. Beyond breaking bad news: how to help patients who suffer. West J Med 1999 Oct;171(4):260-263 [FREE Full text] [Medline: 10578682]
14. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES-a six-step protocol for delivering bad news: application to the patient with cancer. Oncologist 2000;5(4):302-311 [FREE Full text] [doi: 10.1634/theoncologist.5-4-302] [Medline: 10964998]

15. SCOPE: Sharing Concerns with Families. URL: https://www.scope.org.uk/Support/Professional/Sharing-Diagnosis [accessed 2018-12-04] [WebCite Cache ID 74QnSR9rd]
16. Schofield PE, Beeney LJ, Thompson JF, Butow PN, Tattersall MH, Dunn SM. Hearing the bad news of a cancer diagnosis: the Australian melanoma patient's perspective. Ann Oncol 2001 Mar;12(3):365-371. [doi: 10.1023/A:1011100524076] [Medline: 11332150]
17. Baird G, McConachie H, Scrutton D. Parents' perceptions of disclosure of the diagnosis of cerebral palsy. Arch Dis Child 2000 Dec;83(6):475-480 [FREE Full text] [doi: 10.1136/adc.83.6.475] [Medline: 11087279]
18. Garwick AW, Patterson J, Bennett FC, Blum RW. Breaking the news. How families first learn about their child's chronic condition. Arch Pediatr Adolesc Med 1995 Sep;149(9):991-997. [Medline: 7655604]
19. Makoul G. MSJAMA. Communication skills education in medical school and beyond. J Am Med Assoc 2003 Jan 1;289(1):93. [Medline: 12503986]
20. Makoul G. The SEGUE Framework for teaching and assessing communication skills. Patient Educ Couns 2001 Oct;45(1):23-34. [Medline: 11602365]
21. Jonathan S, Suzanne K, Juliet D. Skills For Communicating With Patients. Abingdon, Oxon, UK: Radcliffe Publishing; 1998.
22. Seifart C, Hofmann M, Bär T, Riera Knorrenschild J, Seifart U, Rief W. Breaking bad news-what patients want and what they get: evaluating the SPIKES protocol in Germany. Ann Oncol 2014 Mar;25(3):707-711 [FREE Full text] [doi: 10.1093/annonc/mdt582] [Medline: 24504443]
23. Abazari P, Taleghani F, Hematti S, Ehsani M. Exploring perceptions and preferences of patients, families, physicians, and nurses regarding cancer disclosure: a descriptive qualitative study. Support Care Cancer 2016 Dec;24(11):4651-4659. [doi: 10.1007/s00520-016-3308-x] [Medline: 27296237]
24. Kreuter MW, Sugg-Skinner C, Holt CL, Clark EM, Haire-Joshu D, Fu Q, et al. Cultural tailoring for mammography and fruit and vegetable intake among low-income African-American women in urban public health centers. Prev Med 2005 Jul;41(1):53-62. [doi: 10.1016/j.ypmed.2004.10.013] [Medline: 15916993]
25. Parker PA, Baile WF, de Moor C, Lenzi R, Kudelka AP, Cohen L. Breaking bad news about cancer: patients' preferences for communication. J Clin Oncol 2001 Apr 1;19(7):2049-2056. [doi: 10.1200/JCO.2001.19.7.2049] [Medline: 11283138]
26. Choudhry A, Hong J, Chong K, Jiang B, Hartman R, Chu E, et al. Patients' preferences for biopsy result notification in an era of electronic messaging methods. JAMA Dermatol 2015 May;151(5):513-521. [doi: 10.1001/jamadermatol.2014.5634] [Medline: 25831475]
27. Intille SS. A new research challenge: persuasive technology to motivate healthy aging. IEEE Trans Inf Technol Biomed 2004 Sep;8(3):235-237. [doi: 10.1109/TITB.2004.835531] [Medline: 15484427]
28. Kay M, Morris D, schraefel MC, Kientz J. There's no such thing as gaining a pound: Reconsidering the bathroom scale user interface. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing.: ACM; 2013 Presented at: UbiComp 2013; September 8-13, 2013; Zurich, Switzerland p. 401-410. [doi: 10.1145/2493432.2493456]
29. Chung C, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. CSCW Conf Comput Support Coop Work 2016 Dec 27;2016:770-786 [FREE Full text] [doi: 10.1145/2818048.2819926] [Medline: 28516171]
30. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: an interview study. AMIA Annu Symp Proc 2016;2016:1303-1312 [FREE Full text] [Medline: 28269928]
31. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
32. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
33. Davis GT, Singh H. Should patients get direct access to their laboratory test results? An answer with many questions. J Am Med Assoc 2011 Dec 14;306(22):2502-2503. [doi: 10.1001/jama.2011.1797] [Medline: 22122864]
34. Baldwin DM, Quintela J, Duclos C, Staton EW, Pace WD. Patient preferences for notification of normal laboratory test results: a report from the ASIPS Collaborative. BMC Fam Pract 2005 Mar 8;6(1):11 [FREE Full text] [doi: 10.1186/1471-2296-6-11] [Medline: 15755328]
35. Christensen K, Oldenburg J. Giving patients their results online might be the answer. Arch Intern Med 2009 Oct 26;169(19):1816; author reply 1816-1816; author reply 1817. [doi: 10.1001/archinternmed.2009.369] [Medline: 19858449]
36. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
37. Calvo R, D'Mello S, Gratch J, Kappas A. The Oxford Handbook of Affective Computing. Oxford: Oxford Library of Psychology; 2015.
38. Luneski A, Konstantinidis E, Bamidis PD. Affective medicine. A review of affective computing efforts in medical informatics. Methods Inf Med 2010;49(3):207-218. [doi: 10.3414/ME0617] [Medline: 20411209]

39. Picard RW. Affective computing: challenges. Int J Hum Comput Stud 2003 Jul;59(1-2):55-64. [doi: 10.1016/S1071-5819(03)00052-1]
40. Klein J, Moon Y, Picard R. This computer responds to user frustration: theory, design, and results. Interact Comput 2002 Feb;14(2):119-140. [doi: 10.1016/S0953-5438(01)00053-4]
41. Bickmore T, Pfeifer L, Jack B. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the SIGCHI conference on human factors in computing systems.: ACM; 2009 Apr Presented at: CHI'09; April 04 - 09, 2009; Boston, MA p. 1265-1274. [doi: 10.1145/1518701.1518891]
42. Prendinger H, Dohi H, Wang H, Mayer S, Ishizuka M. Empathic embodied interfaces: Addressing users' affective state. : Springer; 2004 Presented at: Tutorial and Research Workshop on Affective Dialogue Systems; June 14-16, 2004; Kloster Irsee, Germany p. 53-64. [doi: 10.1007/978-3-540-24842-2_6]
43. Legaspi R, Kurihara S, Fukui K, Moriyama K, Numao M. An empathy learning problem for HSI: To be empathic, self-improving and ambient. 2008 Presented at: Conference on Human system interactions; May 25-27, 2008; Krakow, Poland p. 209-214. [doi: 10.1109/HSI.2008.4581435]
44. Boehner K, DePaula R, Dourish P, Sengers P. Affect: from information to interaction. In: Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility. 2005 Presented at: CC'05; August 20-24, 2005; Aarhus, Denmark p. 59-68. [doi: 10.1145/1094562.1094570]
45. Kim MI, Johnson KB. Personal health records: evaluation of functionality and utility. J Am Med Inform Assoc 2002;9(2):171-180 [FREE Full text] [Medline: 11861632]
46. Liu L, Shih P, Hayes G. Barriers to the adoption and use of personal health record systems. In: Proceedings of the 2011 iConference. 2011 Presented at: iConference'11; February 8 -11, 2011; Seattle, Washington, USA p. 363-370. [doi: 10.1145/1940761.1940811]
47. Consolvo S, Roessler P, Shelton B, Lamarca A, Schilit B, Bly S. Technology for care networks of elders. IEEE Pervasive Comput 2004;3(2):22-29. [doi: 10.1109/MPRV.2004.1316814]
48. Suh H, Shahriaree N, Hekler E, Kientz J. The User Burden Scale: A Tool for Assessing User Burden in Computing Systems. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York: ACM; 2016 Presented at: CHI'16; May 7–12, 2016; San Jose, California p. 3988-3999. [doi: 10.1145/2858036.2858448]
49. Kientz J, Choe E, Birch B, Maharaj R, Fonville A, Glasson C, et al. Heuristic evaluation of persuasive health technologies. In: Proceedings of the 1st ACM International Health Informatics Symposium. 2010 Presented at: IHI'10; November 11 - 12, 2010; Arlington, Virginia, USA p. 555-564. [doi: 10.1145/1882992.1883084]
50. Tang C, Carpendale S. Evaluating the deployment of a mobile technology in a hospital ward. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work. 2008 Presented at: CSCW'08; November 08 - 12, 2008; San Diego, CA p. 205-214. [doi: 10.1145/1460563.1460596]
51. Toscos T, Connelly K, Rogers Y. Best intentions: Health monitoring technology and children. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5-10, 2012; Austin, Texas p. 1431-1440. [doi: 10.1145/2207676.2208603]
52. Civan A, Pratt W. Threading together patient expertise. AMIA Annu Symp Proc 2007 Oct 11:140-144 [FREE Full text] [Medline: 18693814]
53. Preece J. Empathy online. Virtual Reality 1999 Mar;4(1):74-84. [doi: 10.1007/BF01434996]
54. Pfeil U, Zaphiris. Patterns of empathy in online communication. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI'07; April 28-May 3, 2007; San Jose, California, USA p. 919-928. [doi: 10.1145/1240624.1240763]
55. Grassi L, Giraldi T, Messina EG, Magnani K, Valle E, Cartei G. Physicians' attitudes to and problems with truth-telling to cancer patients. Support Care Cancer 2000 Jan;8(1):40-45. [doi: 10.1007/s005209900067] [Medline: 10650896]
56. Orlander JD, Fincke BG, Hermanns D, Johnson GA. Medical residents' first clearly remembered experiences of giving bad news. J Gen Intern Med 2002 Nov;17(11):825-831 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10915.x] [Medline: 12406353]
57. Fallowfield L, Jenkins V. Communicating sad, bad, and difficult news in medicine. Lancet 2004 Jan 24;363(9405):312-319. [doi: 10.1016/S0140-6736(03)15392-5] [Medline: 14751707]
58. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
59. TAMS Analyzer for Macintosh OS X.: OA Librarian; 2018. URL: http://tamsys.sourceforge.net/ [accessed 2018-12-05] [WebCite Cache ID 74QnHnroM]
60. Buckman R, Kason Y. How to Break Bad News: A Guide for Health Care Professionals. Baltimore: Johns Hopkins University Press; 1992.
61. Kubler-Ross E, Kessler D. On Grief and Grieving: Finding the Meaning of Grief Through the Five Stages of Loss. New York: Simon & Schuster Adult Publishing Group; 2005.
62. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]

63. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
64. Hekler E, Klasnja P, Froehlich J, Buman M. Mind the theoretical gap: interpreting, using,developing behavioral theory in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France p. 3307-3316. [doi: 10.1145/2470654.2466452]
65. 23andMe. URL: http://23andme.com [accessed 2018-12-05] [WebCite Cache ID 74Qo3YgpQ]
66. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med 2008 May;34(5):373-381 [FREE Full text] [doi: 10.1016/j.amepre.2007.12.024] [Medline: 18407003]
67. Strecher VJ, Shiffman S, West R. Moderators and mediators of a web-based computer-tailored smoking cessation program among nicotine patch users. Nicotine Tob Res 2006 Dec;8(Suppl 1):S95-101. [Medline: 17491176]
68. Kreuter MW, Bull FC, Clark EM, Oswald DL. Understanding how people process health information: a comparison of tailored and nontailored weight-loss materials. Health Psychol 1999 Sep;18(5):487-494. [Medline: 10519465]
69. Suh H, Porter J, Hiniker A, Kientz JA. @BabySteps: design and evaluation of a system for using twitter for tracking children's developmental milestones. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems. New York: ACM; 2014 Presented at: CHI'14; April 26-May 1, 2014; Toronto, Canada p. 2279-2288. [doi: 10.1145/2556288.2557386]
70. Suh H, Porter JR, Racadio R, Sung Y, Kientz JA. Baby steps text: feasibility study of an SMS-based tool for tracking children's developmental progress. AMIA Annu Symp Proc 2016;2016:1997-2006 [FREE Full text] [Medline: 28269959]
71. Jenni OG, Chaouch A, Caflisch J, Rousson V. Infant motor milestones: poor predictive value for outcome of healthy children. Acta Paediatr 2013 Apr;102(4):e181-e184. [doi: 10.1111/apa.12129] [Medline: 23289493]
72. Murray GK, Jones PB, Kuh D, Richards M. Infant developmental milestones and subsequent cognitive function. Ann Neurol 2007 Aug;62(2):128-136 [FREE Full text] [doi: 10.1002/ana.21120] [Medline: 17487877]
73. Friedman B, Kahn PHJR. Human values, ethics, and design. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Mahwah, NJ: Lawrence Erlbaum Associates; 2003:1177-1201.
74. Wright P, McCarthy J. Empathy and experience in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI'08; April 5-10, 2008; Florence, Italy p. 637-646. [doi: 10.1145/1357054.1357156]

Abbreviations

EMRs: electronic medical records
PHRs: personal health records

Edited by G Eysenbach; submitted 01.09.17; peer-reviewed by C Bond, D Hansen, G Petrič, E Murnane, E Eikey; comments to author 02.12.17; revised version received 05.12.18; accepted 07.02.19; published 17.05.19.

Please cite as:
Choe EK, Duarte ME, Suh H, Pratt W, Kientz JA
Communicating Bad News: Insights for the Design of Consumer Health Technologies
JMIR Hum Factors 2019;6(2):e8885
URL: http://humanfactors.jmir.org/2019/2/e8885/
doi:10.2196/humanfactors.8885
PMID:31102374

Original Paper

Complementing a Clinical Trial With Human-Computer Interaction: Patients' User Experience With Telehealth

Sakib Jalil[1], PhD; Trina Myers[2], PhD; Ian Atkinson[3], PhD; Muriel Soden[4], MPH, MD

[1]Betty Irene Moore School of Nursing, UCDavis Health, University of California - Davis, Sacramento, CA, United States

[2]Information Technology, James Cook University, Townsville, Australia

[3]eResearch Centre, James Cook University, Townsville, Australia

[4]School of Medicine, James Cook University, Townsville, Australia

Corresponding Author:
Sakib Jalil, PhD
Betty Irene Moore School of Nursing
UCDavis Health
University of California - Davis
2450 48th Street
Sacramento, CA, 95817
United States
Phone: 1 2139158762
Email: dr.sakib.jalil@gmail.com

Abstract

Background: The use of telehealth to monitor patients from home is on the rise. Telehealth technology is evaluated in a clinical trial with measures of health outcomes and cost-effectiveness. However, what happens between a technology and the patients is not investigated during a clinical trial—the telehealth technology remains as a "black box." Meanwhile, three decades of research in the discipline of human-computer interaction (HCI) presents design, implementation, and evaluation of technologies with a primary emphasis on users. HCI research has exposed the importance of user experience (UX) as an essential part of technology development and evaluation.

Objective: This research investigates the UX of patients with type 2 diabetes mellitus (T2D) with a telehealth in-home monitoring device to manage T2D from home. We investigate how the UX during a clinical trial can be researched and what a clinical trial can learn from HCI research.

Methods: We adopted an ethnographic philosophy and conducted a contextual inquiry due to time limitations followed by semistructured interviews of 9 T2D patients. We defined the method as Clinical User-experience Evaluation (CUE). The patients were enrolled in a telehealth clinical trial of T2D; however, this research was an independent study conducted by information technologists and health researchers for a user-centered evaluation of telehealth.

Results: Key analytical findings were that patients valued the benefits of in-home monitoring, but the current device did not possess all functionalities that patients wanted. The results include patients' experiences and emotions while using the device, patients' perceived benefits of the device, and how patients domesticated the device. Further analysis showed the influence of the device on patients' awareness, family involvement, and design implications for telehealth for T2D.

Conclusions: HCI could complement telehealth clinical trials and uncover knowledge about T2D patients' UX and future design implications. Through HCI we can look into the "black box" phenomenon of clinical trials and create patient-centered telehealth solutions.

(JMIR Hum Factors 2019;6(2):e9481) doi:10.2196/humanfactors.9481

KEYWORDS

clinical user-experience evaluation; telehealth; type 2 diabetes; user experience; human-computer interaction; patient-centered; patient-technology interaction; eHealth

Introduction

Background

Type 2 diabetes mellitus (T2D) is currently one of the world's fastest-growing diseases; the prevalence of T2D rose from 171 million persons affected in 2000 to 415 million in 2015 worldwide [1]. The total annual global health expenditure for diabetes in 2015 was US $673 billion. The cost accounted for 12% of the world's total health expenditure [1].

Treatments for T2D involve diet control, exercise, home blood glucose testing, and, in some cases, oral medication with or without insulin [2]. Effective individualized treatments may also incorporate psychosocial, lifestyle, and other medical interventions [3].

Technology-mediated treatments, such as telehealth, eHealth, mHealth to monitor patients from their homes, are on the increase with chronic diseases such as T2D. Telehealth is the use of information and communication technology (ICT) to provide clinical treatments over distances [4]. A common telehealth treatment for T2D is for patients to send regular blood glucose data to nurses or health care providers via phone, tablet, computer, Web-based system, videoconference, phone call, or short message service (SMS) text [5,6]. A nurse or health care provider is involved in T2D telehealth treatments continuously, while the technology intervention remains as a means of transferring data (eg, blood glucose, blood pressure) and facilitates the communication between patients and nurses for better management of T2D [5,6].

During evaluation through randomized clinical trials, telehealth technology is represented as a "black box." Systematic reviews have shown that clinical trials assess "what went in" (eg, baseline measures) and "what came out" (eg, postintervention measures). "What happens inside the interventions" (eg, how patients felt about using the device and the development of the interventions not achieving a match between technology and context) is rarely a focus of attention in clinical trials [7,8]. For example, in a clinical trial of T2D, the long-term blood glucose HbA1c of patients at baseline is compared against HbA1c at the end of the trial. Improvements in HbA1c, along with additional health parameters, are data that the clinical researchers use to conclude whether a telehealth technology for T2D was effective or not.

Clinical trials do not investigate the relationship between the technology and effects of the use on patients as technology users, how patients interact with these technologies, or how patients feel when using these technologies [5,8]. However, the discipline of human-computer interaction (HCI) tends to be highly divergent in the choice of methods and approaches to understand humans and their interactions. A common practice in HCI is to understand user experience (UX) to design and develop a human-centered technology. UX refers to how a product behaves and is used by people in the real world [9].

We were interested in solving the "black box" phenomenon of a telehealth T2D clinical trial. We looked at six common methods (Table 1) of HCI to explore if we could use one or more of them during clinical trials to understand the UX of patients with T2D with telehealth.

Upon investigation of the six methods in Table 1, we concluded that there was no possibility to conduct a codesign, participatory design, lead user approach, or empathic design because these methods are conducted to create new solutions along with stakeholders. In a clinical trial, a device already in use is selected already by doctors, nurses, and stakeholders. Next, the effectiveness of the device is evaluated, and user-centered design methods are not practiced in a clinical trial. Therefore, we were only left with two options: applied ethnography and contextual design inspired by ethnography.

We adopted an ethnographic philosophy for this study to understand how the situation is in a clinical trial by moving the researchers into the users' environment. Due to time and resource restrictions, we deduced to conduct a contextual inquiry and observations, followed by a semistructured interview, and finally another follow-up via survey. This HCI-inspired research method was named Clinical User-experience Evaluation (CUE) [6]. We wanted to conduct an independent study from an HCI perspective; therefore, we went through a process of defining CUE and its differences from the clinical trial. This paper presents the results of the UX of patients with T2D with telehealth.

Table 1. The 6 dominant human-computer interaction methods.

Method	Key feature	Research orientation
Applied ethnography [10]	Long-term immersive fieldwork; observation combined with participation	Researcher moves into users' world
Contextual design [11]	An ethnographic approach to finding the specific needs of users in a work situation; provides 8 methodological steps	Researcher moves into users' world
Empathic design [12]	Draws on information about the user and her everyday life, and includes inspiration for design and empathy, or "a feel" for the user	Researcher moves into users' world
Participatory design [13]	Users who will be using a system are given a role in the design, evaluation, and implementation of the system	Users brought into the researcher's world
Co-design [14]	May invite users and other people who do not yet know each other; design a product for a mass market or nonwork contexts	Users brought into the researcher's world
Lead user approach [15]	Brings innovative users together, as many ideas of new products or services originate in the minds and hands of them and not from professional researchers and designers	Users brought into the researcher's world

Research Objective

The research objective was to investigate how to discover patients' UX in telehealth, eHealth, and mHealth in a clinical trial. To pursue the research objective, we answered the following three questions with the CUE:

1. What happens at the patient's home during the use of the telehealth device?
2. How do patients feel while using the telehealth device?
3. Which function(s) and designs of the device satisfies/dissatisfies the patients?

Methods

Research Method

An investigation through meta-synthesis conducted in 2014 of past clinical trials of telehealth T2D concluded that there is a need for new practices that could capture the experience of users (patients) in a clinical trial [6]. Therefore, we created the CUE. The CUE consisted of three stages (Figure 1). Stage one was a contextual inquiry performed in situ at a patient's home. During this stage, a patient used the device with the think-aloud method as one researcher as the observer took notes. This contextual inquiry was conducted during a patient's regularly scheduled time for using the device, in the patient's home. Stage two was a semistructured qualitative inquiry into the patients' experience and expectations, the questions that developed during stage one, and anything extra the patient wanted to talk about. The interview took place directly after stage one on the same day, while perceptions were still fresh in the mind of the user. Stage three was an anonymous survey to follow-up with patients the findings from the first two stages and if there were any changes in the use of the device. This was conducted 8 months after stage two. The researchers were ICT researchers from James Cook University Townsville (Queensland, Australia) who had no involvement with the clinical trial. Every participant was enrolled at least 3 months (12 weeks) into the clinical trial to avoid novelty effects.

During the application of the CUE, health professionals asked us (the HCI researchers) to articulate the contribution of CUE as opposed to a clinical trial, especially because the clinical trial is a 300-year-old methodology [16] used in medical science. The CUE protocol is compared to the clinical trial in Table 2 to show the differences. Because clinical trials are regulated protocols, this table supported us to convey the information to the team of health scientists (nurses and physicians).

Figure 1. The Clinical User-experience Evaluation (CUE) methodology.

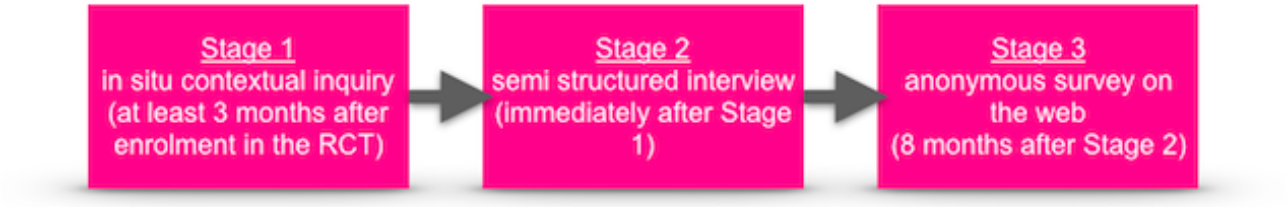


Table 2. Differences between the Clinical User-experience Evaluation (CUE) and clinical trials.

Review criteria	CUE[a]	Clinical trial
Investigation aims	Investigates patients' experience, understanding, feeling, and usage of a technology for health care	Investigates patients' medical condition with an intervention that can be a drug or a technology
Outcome	To provide patient feedback about using the trial technologies and a guide for future improvement of the technology, including features that were lacking or nonexistent that would benefit the treatment process	To provide enough evidence for medical practitioners to make sound judgments
Sample size	A smaller sample population similar to HCI[b] qualitative user evaluation is appropriate	Requires large sample population to provide substantial and robust evidence
Regulations	Tests interaction with a device without interfering in any medical protocols, there is no physical or psychological stress; conducted at the regular times a patient uses the technology as part of the overarching clinical trial	Rigorous form of testing that must follow HTA[c] guidelines; clinical trials often include psychosocial analysis questionnaire
Investigator	Can be carried out by anyone working in the field of HCI with simple practice and observational skills	Carried out by medical staff or caregivers who have either medical credentials or training in health care and/or social work
Recruitment	Participants come from the clinical trial	Larger samples of volunteers are sought who have specific medical conditions
Ethics	Privacy of information is required, and the participant must provide written consent	Strong, regulated ethical process and abiding by HTA regulations

[a]CUE: Clinical User-experience Evaluation.

[b]HCI: human-computer interaction.

[c]HTA: Health Technology Assessment.

Participants: Inclusion Criteria

The CUE was applied on a clinical trial that was conducted by Townsville-Mackay Medicare Locals in North Queensland, Australia [17]. A total of 210 patients were recruited in Townsville, Mackay, and Brisbane. Participants were referred by two nurses. The participants of the CUE were (1) enrolled in the clinical trial, (2) belonged in the intervention group (using the telehealth device), (3) diagnosed with T2D for at least 12 months, and (4) volunteered to participate in CUE.

Participant Details

Participation in the CUE was voluntary. A total of 12 patients initially agreed to participate. However, three of them opted out of the CUE study because they were not available during the designated time frame. Nine patients participated in the CUE study. Five of them were considered part of the aging population with an age of at least 64 years, and four participants were within the age range of 50 to 63 years (Table 3). Participants were given pseudonyms that were incredibly different from the participants' original names. In addition to the nine participants, five family members occasionally provided feedback. Of these five, only two family members permitted us to use their quotes.

Table 3. Participant details (N=9).

Participants pseudonyms	Sex	Age (years)	Computer use (hours/week)	Time in clinical trial (months)	Time since diagnosed with T2D (years)
Uma	Female	74	0	5	>12
Zach	Male	70	70	8	>10
Yanicka	Female	68	20	6	7
Vince	Male	66	20	6	>10
Bill	Male	64	4	5	20
Heidi	Female	60	2	5	25
Serena	Female	55	12	3	2
Pete	Male	53	2	6	1
Ted	Male	52	60	6	2

Figure 2. The in-home monitoring device of the clinical trial: a tablet PC, sphygmomanometer, and glucometer.

Equipment

Participants used a tablet computer with an 11-inch screen, an automatic glucometer, and an automatic sphygmomanometer (Figure 2). The device had a touchscreen interface and was a single-user system. A regular patient session entailed a patient turning on the tablet and waiting to log in automatically. The patient then looked at the scheduled blood glucose and blood pressure test that was arranged by the nurse. The patient pricked a finger to get a drop of blood and put it on a strip for a blood glucose reading. The strip was then placed in the glucometer. To get the blood pressure measurement, the patient put an arm in the sphygmomanometer, which automatically took the reading.

Data Analysis

Interviews and contextual inquiry sessions were audio recorded. The recordings were transcribed, and the notes and data from the contextual inquiry were analyzed using the contextual design methodology (Multimedia Appendix 1). The semistructured interviews were analyzed with thematic content analysis; NVivo 10 software was used to manage the analysis process.

Results

The results showed two themes: (1) the current design and how that fits with the patients' needs, and (2) the patients' experience

of using the device depicted through their feelings and perceptions.

Current Design: Technology-User Fit

Placement of the Device

We found that patients placed the device in different parts of their homes (Table 4). The patients chose to place the device in four locations: living room (n=4), bedroom (n=2), study room (n=2), and patio (n=1). Reasons mentioned were internet or phone socket availability (n=3), convenience (n=4), comfort (n=1), and self-motivation (n=1) regarding their choice of device placement.

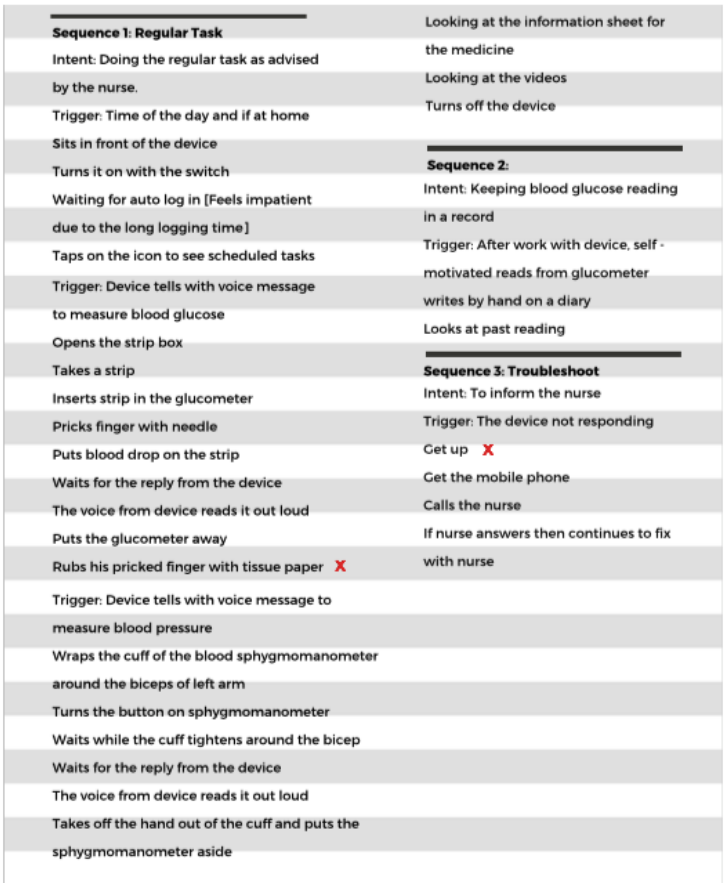
Data from the contextual inquiry was first analyzed through four steps of the contextual design method (see Multimedia Appendix 1). The results from one exemplary case, Zach's sequence model, showed that he had three breaks noted with a red mark (Figure 3). The breaks were (1) to save data because previous readings were not saved for the patients, only for the nurses, (2) to clean his fingertips after the blood test to continue with the touchscreen, and (3) to use the internet on a different device because the telehealth device did not have names of all medications.

Table 4. Placement of the device in the patients' homes (N=9).

Reason	Location, n				Total for reason, n
	Living room	Study room	Bedroom	Patio	
Internet socket	2	1	—[a]	—	3
Comfort	1	—	—	—	1
Convenience	1	1	1	1	4
Self-motivation	—	—	1	—	1
Total in each room	4	2	2	1	

[a]Room-reason not selected.

Figure 3. Example of one sequence diagram that shows breaks of patient.



Lack of Wireless Capability

The device only functioned with wired internet that had to be connected through a cable through the telephone port in a patient's house. Heidi, Serena, and Uma mentioned that having wires was a problem of the device:

> *Apart from when you gotta be home two hours after eating to do it can be a little bit difficult like, "Oh my God I have gotta get home," so, I mean, time-wise that's it if I am not gonna be at home.* [Heidi]

When Uma traveled, she used a separate glucometer and would keep her blood glucose readings in a diary. She would later come home and update her nurses about the data. However, the device did not allow users to record data manually:

> *I can take this [her own glucometer that she bought] with me, I can't do the blood pressure, I take this with me and do the blood sugar and then put it down in a book.* [Uma]

Undesirable Experience From Sphygmomanometer

Every patient criticized the sphygmomanometer. It was difficult to use. It also gave uncomfortable experiences:

> *The blood pressure cuff I have more difficulty with. I put it here where my doctor would put it. It repumps, and it takes ages to do it. It marks my arm.* [Yanicka]

Yanicka complained of physical pain around her arm from the device. She stated that this pain was more than other sphygmomanometers that she had used in doctor visits.

Lack of Visual Data

In the current system, each time the patients conducted a test, they were presented with instant data on their blood pressure and blood glucose levels. However, when the patients conducted the next scheduled test, they could not see the previous data. For example, if a patient did a test in the morning and one in the evening, they were unable to compare the readings, because the earlier test was not available. Patients expressed their desire and the importance to see the previous data to help them know if they were doing better or worse in terms of their blood glucose:

> *I know it does it here [glucometer], but it would be good to see every day's. But it doesn´t show you. Like last week I might have been 5.5 and this week I am 7.5. Why? Why am I? Then I would do exactly the same things that I did last week.* [Bill]

Vince and his wife also mentioned the adjustment of insulin, similar to Heidi. They said that while Vince took insulin and was adjusting the dosage of the insulin, they would prefer to see a day-by-day comparison of Vince's blood sugar in a graph:

> *It would be much better if he could just push a button and see the last three weeks of his readings.*
>
> *Coz he is adjusting his insulin and he needs to know—all the time.* [Vince's wife]
>
> *Probably I would like to see a graph of my results, more often. Like even once a month would be good to show it on a graph. How my results are going, because you just see number every day, but you want to know your ups and downs, and you want to know using that computer why my diabetes goes higher, I know the reason now why it goes higher, before I didn´t know the reasons. But now I do. And it's just the difference the food that I have eaten, and the foods prepared, and I have found that because I am monitoring my blood glucose carefully.* [Pete]

Zach stated that graphs are a great tool to compare trends. Zach was very particular about using a progressive graph. He also commented that much research is required on how to show the blood pressure and the blood glucose level in the graph:

> *There is nothing like graphs to see trends. They have to display in a sensible way, if that makes sense. I will be thinking that a progressive graph will do it.* [Zach]

Lack of Medication Name

Yanicka stated that the medication that she was taking was not included in the information sheet listed on the device. This meant that the database did not contain a full list of all possible diabetes medications that the patients in this clinical trial were using. This necessitated Yanicka using another computer to locate information about the medication that was prescribed for her:

> *To see my change of insulin and I couldn't find on here, so I went back through here with my computer and internet. My medicine is also here...and insulin is not there, but I looked that up at the computer. Not everyone has that. When I want to see what that thing do I check it up here. I don't ever touch the unit because it automatically shuts down. It's simple as that, quite easy to use. Bit challenging at the beginning.* [Yanicka]

Zach reported the same problem—his medication was missing from the available information sheet on the device.

Mismatch With Life Due to Immobility of the Device

The device currently works only with internet cables. All the patients stated that a mobile unit would have been much more suitable than the current device. Uma stated that she could not carry the device. So she carried a different glucometer to keep the data for her records:

> *I can take this with me; I can't do the blood pressure, I take this with me and do the blood sugar and then put it down in a book.* [Uma]

Glucometer Discomfort and Pain

When a patient uses a glucometer, a small drop of blood is obtained by pricking the skin with a lancet. The drop of blood is placed on a disposable test strip that the meter reads and uses to calculate the blood glucose level. Slight discomfort is experienced when the lancet pricks the skin of the finger. However, T2D patients use a glucometer frequently, often more than once a day. Some of the patients in this clinical trial mentioned the discomfort and pain from the glucometer. Ted stated that after frequent use over a long period, his finger feels bruised:

> *Problem I see with this is you have to prick your finger every time you use it. It's not that bad but after a while you are bruising your fingertips sore, so in that respect I guess it's not really something that one looks forwards to going and doing.* [Ted]

Every other patient also felt the pain and complained of being hurt. As a remedy, Zach was interested to see what the scientists come up with in the future. Ted also mentioned that he wants science to advance in such a way that a chip can be inserted and left in a human body so that it will transmit continuous readings to the machine. In this way, Ted thinks, bruising and pain may be avoided.

Feelings and Perceptions

Patients used words such as "motivation," "accountability," "safety net," "habit," and "awareness" while they expressed their frustrations with the telehealth device.

Motivation

Two participants, Vince and Heidi, mentioned that using the device motivated them to manage their diabetes:

> *And it's good that they [nurses], that someone else is keeping an eye on you, back at office, nurses.* [Vince]
>
> *And it gives you just that extra push, you know?* [Heidi]

Build a Habit

Pete lives alone, and he stated that he had developed a habit from using the device for 6 months in the clinical trial. His habit was measuring his blood glucose and blood pressure early in the morning before he would engage in his daily life:

> *I think it's a great benefit for me, I wish it probably could stay, and I would like to keep it. I don´t know how I am gonna go; I am obviously in the habit of doing it every morning now, I am gonna have it. It´s a habit now. So next week it's gonna go, and I can still maintain the regimen that I am doing it now, you know.* [Pete]

Awareness

Enrollment in the clinical trial had made Serena aware of her well-being. The device would make her do things regularly. Serena called this being in a regimen where she had to regularly monitor and be aware of her blood glucose and her food. Serena's son, who was one of the family members to permit his data to be used in this research, mentioned:

> *It's more like a—there's a regimen for every day 10 minutes before eating and after eating, she tastes it and morning, afternoon—it's 10 minutes or 5 minutes—doesn't affect much. But it improved her overall awareness.* [Serena's son]

Vince stated that after he had looked at the results, he felt more aware and accountable, which made him want to use the device more:

> *It [the device] makes you, wanna do it [the blood glucose reading].* [Vince]

Heidi compared the use of the telehealth device with quitting smoking. In quit-smoking programs, people are typically encouraged to call a back-end, or a buddy, each time they have the urge to smoke. Heidi found using the device a similar experience as it makes her do the one extra step that she needs to take:

> *You know when you haven't done this for a week, and oh you should do it. It's like quitting smoking; you know that you have to ring up somebody every time you have to ring up. So, it's that extra incentive you know.* [Heidi]

Feel Safe

Daily monitoring provided safety and comfort to the patients. In the case of Vince, daily monitoring made his wife feel safe that someone was watching over him:

> *It's sort of like a safety net. You know there's someone in the background always watching and they will ring you up.* [Vince's wife]

For Uma (a 74-year-old woman living alone), the device was not of interest. In Uma's opinion, the use of the device provided the nurses with the data that they needed and that made her feel safe. Serena's son stated that Serena's enrollment in the trial and use of the device helped him to look after her.

Reduced Doctor Visits

Patients stated that they had fewer visits to the doctor during the time enrolled in the clinical trial. They indicated that they did not have to see a doctor every 3 months, which is the traditional treatment. Instead, they spoke with the nurse every 2 weeks, which decreased the doctor visits unless there was something urgent.

Frustrations

Patients had frustrations using the device due to slow responses and sometimes during unresponsive states. Even after participating in the clinical trial for more than 3 months, the patients often had problems with the device. For example, 74-year-old Uma, in her fifth month in the clinical trial, was very frustrated during the contextual inquiry. A portion of the transcript (from the second minute until the fifth minute) of Uma is as follows:

> *Uma: I don't know what's wrong with it; it suddenly slowed down.*
>
> *Researcher: Did it slow down today or—?*
>
> *Uma: No, it has been doing this for a few days. I was talking to the lady [Nurse1] on the phone and—come on.*

Uma called "come on" to the device after being frustrated with the device for not responding to her touches.

> *Uma: I have to go through this every morning. It's—aaah.*

Uma ceaselessly showed frustration, sighed heavily with hand gestures toward the device, and talked to the device.

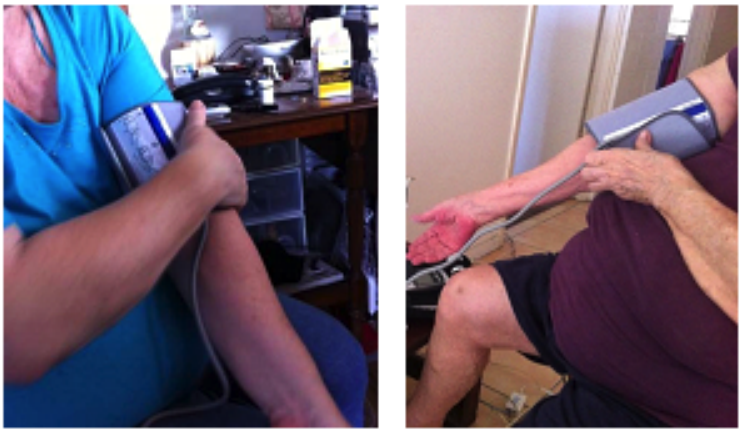
> *Uma: I don't know whether it's because it's—aaaahhhhh.* [more frustration]

After the fifth minute, Uma was able to use the device after restarting it and being helped by the researcher.

Difficulty in Measurement of Blood Pressure

All but one patient (Ted) complained about difficulty with the automatic sphygmomanometer because they had to use one arm to put the cuff around the other arm and then press a button on the device screen to start the automatic adjustment process (Figure 4). It was a very difficult process for any person to do this task alone. Heidi described it as: "It's not really a one-man job."

Figure 4. Heidi (left) and Uma (right) struggling with the sphygmomanometer.



Discussion

Overview

All the patients were more than 3 months in the clinical trial. Yet, we saw frustrations during use of the device due to design—the responses and limitations. There were perceived benefits and promises if designed right. Even how the treatment was designed was influencing patients UX. For example, T2D patients had to measure their own blood pressure, which is not an easy task. Even the researchers could not measure blood pressure accurately with the same device during some practices.

Design Implications for Future Telehealth for T2D

The patients wanted to see their own data meaningfully presented through graphs. And a wireless device was preferred due to mobility. Glucometer comfort, inclusion of all medication names, and wireless connectivity are essential for a device for T2D.

Domestication of the Device

The patients treated the device like regular domestic technology. Stable and compelling routines at home influences the use of domestic technologies [18]. Therefore, considerations of people's routine activities and contexts are essential to inform the design; otherwise, people end up excluding those technologies. Our results resemble Crabtree's findings [19]; in domestic settings, the patients might have multiple other gadgets, and the telehealth device became one of them. Ted placed his device in his living room beside his reclining chair, which shows comfort as a reason. Other participants, such as Vince and Yanicka, chose their device locations based on convenience.

Influence on Patients From the Use of the Technology

The study of how to design technology to motivate behavioral change has been of increased interest to researchers and industrial practitioners due to the widespread use of technology, such as computers, mobile phones, iPads, etc. Persuasive technology is "a computing system, device, or application designed to change a person's attitude or behavior in a certain way" without using coercion or deception [20]. Additionally, technology is never neutral; it influences users in one way or another [21,22]. Persuasive technology is designed to target a specific behavioral change of the users intentionally. These study patients mentioned different levels of influence on their lives from the use of the device. Heidi said she received extra motivation from this device to do her regular blood glucose check. Vince felt motivated to manage his blood glucose because the device motivated him to check it. Pete was motivated by placing the device beside his bed. Serena and her son mentioned during the interview that Serena was more accountable to look after her blood glucose while using the telehealth device. Serena's son stated that Serena was more aware of her blood glucose and food intake after using the device. Additionally, Vince, Heidi, Pete, and Ted also mentioned an improvement in awareness.

This telehealth device was not designed to motivate, build habits, or create awareness among patients. But this device did show the potential to change patients' behavior if it had been integrated with persuasive technology strategies [23]. It could be improved by targeting specific behaviors, such as healthy eating habits [24] of T2D patients, to help manage their conditions better.

Categorization of the Patients as Users

All patients did not use the device with the same degree of interest. We found different levels of interest in the patients based on the observations and their explanations during stages 1 and 2 of the CUE. Our persona categorization of the nine patients in the CUE includes enthusiastic, tolerant, indifferent, and resistant patients [6]. These categories need to be validated with a higher population of patients.

Limitations of the Research

The CUE was conducted with a sample size of nine patients. To generalize these findings across the T2D population, future work should include a higher number of patients and expand quantitatively on findings of this research.

Comparison With Prior Work

Most health researchers advocate larger, well-designed, controlled studies to gather evidence [25-27]. However, there is a gap. There are no studies that evaluate the effectiveness of telehealth in daily practice from patients' lives; rather the studies strengthen current evidence [28]. This research is an approach to bridge that gap and increase evaluation of telehealth from a user (patient) perspective through CUE, unlike some recent usability studies with telehealth. For example, a study conducted for patients with T2D showed that usability improvements increased the acceptability by 57%, but studies of this sort are often explored to gather quantitative evidence only. They do not understand patients, unlike the CUE. To our knowledge, many studies conducted qualitative research as a component added onto a clinical trial, but no study has been conducted from ICT researchers from an HCI perspective that looks at telehealth and its impact on patients as users of these technologies. In another study, a 2016 investigation of patients with T2D who dropped out of an eHealth intervention used semistructured interviews to explore the reasons why patients opt out of a telehealth trial [29]. The CUE in this research used both contextual inquiry with semistructured interviews versus just semistructured interviews and uncovered both satisfied and dissatisfied patients [30].

Past qualitative work reported on telehealth-delivered educational interventions [31] and telephone interventions [31] did not improve medical conditions in T2D patients. Studies such as CUE can explore why some interventions worked or did not work. This kind of investigation had never been conducted in a clinical trial from an HCI perspective by ICT researchers. Generally, HCI evaluation is done during the development phase but, in this study, it was conducted in the rollout phase. Although domestication research had been undertaken with technology and users, domestication of a telehealth in-home monitoring device (in this case for T2D) has not been researched in the past until this study.

Another stream of studies took behavior change approaches in T2D management [32]. The CUE aligns more with this line of research. Researchers in the future should explore more in-depth into the role of the technology intervention and T2D management with approaches like CUE.

Conclusions

Investigation of interactions between patients and a technology are critical in telehealth because it affects the overall outcome of a treatment. Disregard for the needs of patients, social and cultural habits, and the complex nature of health care systems results in relatively low impact and uptake of telehealth and eHealth technologies [33]. Some eHealth and telehealth interventions show dropout rates of up to 80% [34,35], but there is little knowledge about the UX-related dropouts. Therefore, we investigated a telehealth clinical trial through the HCI approach and investigated patients' UX in a T2D clinical trial in Northern Queensland. We discovered that patients benefited from using the in-home monitoring device to manage their T2D regarding awareness, motivation, involvement, etc. Patients' negative experiences with the technology—not all the patients engaged with the telehealth device equally—and design recommendations for future T2D telehealth were also found. We urge a global movement to advocate and practice HCI to complement all telehealth clinical trials and understand patients' UX.

Acknowledgments

The authors would like to thank the Townsville-Mackay Medicare Locals for allowing this study to run in conjunction with their T2D randomized controlled trial [18]. Special thanks to the participants of the T2D trial and Dr Dianna Hardy for her support and expert advice.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Contextual Design Process

[PDF File (Adobe PDF File), 312KB - humanfactors_v6i2e9481_app1.pdf]

References

1. International Diabetes Federation. 2017. URL: https://www.idf.org/aboutdiabetes/what-is-diabetes.html [accessed 2019-03-29] [WebCite Cache ID 77FbN8LFd]
2. Diabetes Australia. 2017. URL: https://www.diabetesaustralia.com.au/ [accessed 2019-03-29] [WebCite Cache ID 77FWDpL7d]
3. Benocci M, Bachlin M, Farella E, Roggen D, Benini L, Troster G. Wearable assistant for load monitoring: recognition of on-body load placement from gait alterations. In: Proceedings of the 4th International ICST Conference on Pervasive Computing Technologies for Healthcare. 2010 Mar 22 Presented at: 4th International ICST Conference on Pervasive

Computing Technologies for Healthcare; March 22-25, 2010; Munich p. 1-8 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5472905
4. Kruse C, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
5. Lee SW, Chan CK, Chua SS, Chaiyakunapruk N. Comparative effectiveness of telemedicine strategies on type 2 diabetes management: a systematic review and network meta-analysis. Sci Rep 2017 Oct 04;7(1):12680 [FREE Full text] [doi: 10.1038/s41598-017-12987-z] [Medline: 28978949]
6. Jalil S, Myers T, Atkinson I. A meta-synthesis of behavioral outcomes from telemedicine clinical trials for type 2 diabetes and the Clinical User-Experience Evaluation (CUE). J Med Syst 2015 Mar;39(3):28. [doi: 10.1007/s10916-015-0191-9] [Medline: 25677954]
7. Black A, 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 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
8. Kelders S. Understanding Adherence to Web-Based Interventions. Enschede: Universiteit Twente; 2012.
9. Norman D, Nielsen J. Neilsen Norman Group. The definition of user experience (UX) URL: https://www.nngroup.com/ [accessed 2019-03-30] [WebCite Cache ID 77FcixebS]
10. Suchman L. Making work visible. Commun ACM 1995;38(9):56-64. [doi: 10.1145/223248.223263]
11. Beyer H, Holtzblatt K. Contextual Design: Defining Customer-Centered Systems (Interactive Technologies). Burlington, MA: Morgan Kaufmann; 1997.
12. Postma C, Lauche K, Stappers P. Trialogues: a framework for bridging the gap between people research and design. Compiegne; 2009 Oct Presented at: 4th International Conference on Designing Pleasurable Products and Interfaces, DPPI 2009; Oct 13-16, 2009; Compiegne, France URL: https://repository.tudelft.nl/islandora/object/uuid:a1172a2c-2550-4854-b895-51ddfb197fad?collection=research
13. Spinuzzi C. The methodology of participatory design. Tech Commun 2005;52(2):163-174.
14. Sanders E, Stappers P. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
15. Luthje C, Herstatt C. The Lead User method: an outline of empirical findings and issues for future research. R&D Management 2004 Nov;34(5):553-568. [doi: 10.1111/j.1467-9310.2004.00362.x]
16. Davis SN, Thompson CJ, Brown MD, Home PD, Alberti KG. A comparison of the pharmacokinetics and metabolic effects of human regular and NPH insulin mixtures. Diabetes Res Clin Pract 1991 Aug;13(1-2):107-117. [Medline: 1773708]
17. Carlisle K, Warren R, Scuffham P, Cheffins T. Randomised controlled trial of an in-home monitoring intervention to improve health outcomes for type 2 diabetes: study protocol. Stud Health Technol Inform 2012;182:43-51. [Medline: 23138078]
18. Crabtree A, Rodden T, Tolmie P, Mortier R, Lodge T, Brundell P, et al. House rules: the collaborative nature of policy in domestic networks. Pers Ubiquit Comput 2014 Mar 8;19(1):203-215. [doi: 10.1007/s00779-014-0771-6]
19. Crabtree A. Designing Collaborative Systems: A Practical Guide to Ethnography. London, United Kingdom: Springer-Verlag London; May 12, 2003.
20. Fogg B. Persuasive technology. Ubiquity 2002 Dec 01;2002(December):2. [doi: 10.1145/764008.763957]
21. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. CAIS 2009;24(1):28 [FREE Full text] [doi: 10.17705/1CAIS.02428]
22. Torning K, Oinas-Kukkonen H. Persuasive system design state of the art and future directions. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Apr 26 Presented at: Persuasive '09 4th international conference on persuasive technology; April 26-29, 2009; Claremont, CA p. 30.
23. Jalil S, Orji R. Integrating persuasive technology to telemedical applications for type 2 diabetes. 2016 Presented at: Persuasive 2016: Persuasive Technologies; April 5-7, 2016; Salzberg, Austria.
24. Orji R, Vassileva J, Mandryk R. LunchTime: a slow-casual game for long-term dietary behavior change. Pers Ubiquit Comput 2012 Jul 7;17(6):1211-1221. [doi: 10.1007/s00779-012-0590-6]
25. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
26. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a Web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015 Mar 03;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
27. Hadjiconstantinou M, Byrne J, Bodicoat DH, Robertson N, Eborall H, Khunti K, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016 Dec 21;18(10):e270 [FREE Full text] [doi: 10.2196/jmir.5991] [Medline: 27769955]
28. Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French D, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and

meta-analysis of controlled trials. J Med Internet Res 2016 Apr 19;18(4):e86 [FREE Full text] [doi: 10.2196/jmir.5425] [Medline: 27095386]
29. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 Dec 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
30. Verhoeven F, van Gemert-Pijnen L, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
31. Wens J, Vermeire E, Hearnshaw H, Lindenmeyer A, Biot Y, Van Royen P. Educational interventions aiming at improving adherence to treatment recommendations in type 2 diabetes: A sub-analysis of a systematic review of randomised controlled trials. Diabetes Res Clin Pract 2008 Mar;79(3):377-388. [doi: 10.1016/j.diabres.2007.06.006] [Medline: 17643546]
32. van Vugt M, de Wit M, Cleijne WH, Snoek FJ. Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. J Med Internet Res 2013 Dec 13;15(12):e279 [FREE Full text] [doi: 10.2196/jmir.2800] [Medline: 24334230]
33. Nijland N. Grounding eHealth: Towards a Holistic Framework for Sustainable eHealth Technologies. Enschede: University of Twente; 2011.
34. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
35. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

Abbreviations

CUE: Clinical User-experience Evaluation
HCI: human-computer interaction
ICT: information and communication technology
T2D: type 2 diabetes mellitus
UX: user experience

Edited by G Eysenbach; submitted 23.11.17; peer-reviewed by A Knudsen, M Hadjiconstantinou; comments to author 15.03.18; revised version received 11.08.18; accepted 12.12.18; published 06.06.19.

Please cite as:
Jalil S, Myers T, Atkinson I, Soden M
Complementing a Clinical Trial With Human-Computer Interaction: Patients' User Experience With Telehealth
JMIR Hum Factors 2019;6(2):e9481
URL: http://humanfactors.jmir.org/2019/2/e9481/
doi:10.2196/humanfactors.9481
PMID:31172958

Original Paper

Live Usability Testing of Two Complex Clinical Decision Support Tools: Observational Study

Safiya Richardson[1], MPH, MD; David Feldstein[2], MD; Thomas McGinn[1], MPH, MD; Linda S Park[2], PhD; Sundas Khan[1], MD; Rachel Hess[3], MS, MD; Paul D Smith[2], MD; Rebecca Grochow Mishuris[4], MS, MPH, MD; Lauren McCullagh[1], MPH; Devin Mann[5], MS, MD

[1]Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States

[2]University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

[3]School of Medicine, University of Utah, Salt Lake City, UT, United States

[4]Boston University School of Medicine, Boston, MA, United States

[5]New York University School of Medicine, New York, NY, United States

Corresponding Author:
Safiya Richardson, MPH, MD
Donald and Barbara Zucker School of Medicine at Hofstra/Northwell
500 Hofstra University
Hempstead, NY, 11549
United States
Phone: 1 404 695 7883
Email: srichard12@northwell.edu

Abstract

Background: Potential of the electronic health records (EHR) and clinical decision support (CDS) systems to improve the practice of medicine has been tempered by poor design and the resulting burden they place on providers. CDS is rarely tested in the real clinical environment. As a result, many tools are hard to use, placing strain on providers and resulting in low adoption rates. The existing CDS usability literature relies primarily on expert opinion and provider feedback via survey. This is the first study to evaluate CDS usability and the provider-computer-patient interaction with complex CDS in the real clinical environment.

Objective: This study aimed to further understand the barriers and facilitators of meaningful CDS usage within a real clinical context.

Methods: This qualitative observational study was conducted with 3 primary care providers during 6 patient care sessions. In patients with the chief complaint of sore throat, a CDS tool built with the Centor Score was used to stratify the risk of group A Streptococcus pharyngitis. In patients with a chief complaint of cough or upper respiratory tract infection, a CDS tool built with the Heckerling Rule was used to stratify the risk of pneumonia. During usability testing, all human-computer interactions, including audio and continuous screen capture, were recorded using the Camtasia software. Participants' comments and interactions with the tool during clinical sessions and participant comments during a postsession brief interview were placed into coding categories and analyzed for generalizable themes.

Results: In the 6 encounters observed, primary care providers toggled between addressing either the computer or the patient during the visit. Minimal time was spent listening to the patient without engaging the EHR. Participants mostly used the CDS tool with the patient, asking questions to populate the calculator and discussing the results of the risk assessment; they reported the ability to do this as the major benefit of the tool. All providers were interrupted during their use of the CDS tool by the need to refer to other sections of the chart. In half of the visits, patients' clinical symptoms challenged the applicability of the tool to calculate the risk of bacterial infection. Primary care providers rarely used the incorporated incentives for CDS usage, including progress notes and patient instructions.

Conclusions: Live usability testing of these CDS tools generated insights about their role in the patient-provider interaction. CDS may contribute to the interaction by being simultaneously viewed by the provider and patient. CDS can improve usability and lessen the strain it places on providers by being short, flexible, and customizable to unique provider workflow. A useful component of CDS is being as widely applicable as possible and ensuring that its functions represent the fastest way to perform a particular task.

(JMIR Hum Factors 2019;6(2):e12471) doi:10.2196/12471

KEYWORDS

usability; usability testing; user experience; clinical decision support; health informatics; provider adoption; workflow; live usability; clinical prediction rules

Introduction

Background

The landmark Institute of Medicine report *To Err Is Human*, sparked an increased focus on the prevention of medical errors [1]. Computerized clinical decision support (CDS) aids providers in clinical decision making for individual patients [2] and was proposed as a key tool to improve quality of care by providers, policy makers, experts, and consumers [1,3,4]. In the United States, unprecedented resources were committed to support the adoption and use of electronic health records (EHRs) through the Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009 including incentive payments by the federal government totaling up to US $27 billion over 10 years [5]. EHR adoption in eligible hospitals and practices grew from less than 10% in 2008 to over 80% in 2015 [6]. One of the HITECH requirements, for meaningful use of EHRs, included criteria to implement CDS at every stage.

CDS can improve quality by improving diagnosis, treatment, and preventative care services [7-20], but it now contributes to the increasing complexity of clinical practice. Murphy et al reported primary care doctors receive 77 notifications in the EHR per day [21] and spend nearly 2 hours on the EHR and desk work for every hour of face-to-face time with their patients [22]. Poor EHR usability is a major driver of declining career satisfaction among providers [23]. CDS is almost never tested in real clinical care sessions that have real-time pressure and patient-case complexity. As a result, many tools that appear usable and useful during development and usability testing, are cumbersome within workflow, are poorly adopted, and fail to deliver on their promise of improved care [14].

There is an extensive literature detailing the features of highly usable CDS. The foundational article “Ten Commandments for Effective Clinical Decision Support” specifies the importance of creating CDS that is fast, anticipates provider needs, fits into user workflow, provides a change in practice as opposed to a stop, is simple with few user inputs, and is adaptive [24]. A comprehensive literature review of studies evaluating barriers to and facilitators of CDS usage details similar CDS-specific usability issues including minimal mouse clicks and workflow integration [25]. These works and many others [26-33] are important but primarily based on expert opinion and provider feedback given via surveys, interviews, and simulated usability testing. Few have objectively observed providers during a real clinical session and none has observed the provider interaction with complex CDS.

Objectives

The objective of this study was to further understand the barriers to and facilitators of meaningful CDS tool usage within a real clinical context. Usability testing of 2 CDS tools was conducted as a part of the study “Integrated Clinical Prediction Rules: Bringing Evidence to Diverse Primary Care Settings (iCPR2),” a randomized controlled trial evaluating the tools’ effect on antibiotic ordering [34]. The CDS tools were composed of an alert, a clinical prediction rule (Centor Score and Heckerling Rule) estimating risk of either group A Streptococcus (GAS) pharyngitis or pneumonia, and an automatic order set based on risk.

Methods

This was a qualitative observational study done in January 2017 at the University of Wisconsin-Madison, School of Medicine, a large academic health care center, where the parent study was being conducted. Testing was completed with a convenience sample of 3 volunteer primary care providers during a total of 6 patient care sessions. Inclusion criteria required that participants (1) worked in Family Medicine or Internal Medicine clinics, (2) spent at least half of their time providing clinical care, and (3) were randomized to the intervention arm of the larger Integrated Clinical Prediction Rules: Bringing Evidence to Diverse Primary Care Settings (iCPR2) study with CDS embedded in their EHR system. The sample size was typical for usability studies and was considered sufficient to elicit the vast majority of usability issues [35-37]. The sample size was considered to be 6, for each patient care session, as each was a complex and unique interaction between the patient, provider, and CDS tool. A typical sample size for usability studies is 5.

The 2 CDS tools tested in the parent study used clinical prediction rules to evaluate the risk of GAS pharyngitis in patients presenting with sore throat (the Centor Score) and the risk of pneumonia in patients presenting with cough or upper respiratory tract infection (the Heckerling Rule). The tools were both built in the EpicCare ambulatory EHR (Epic Corp. Verona, Wisconsin). The tools were triggered by a reason for visit of sore throat, cough, or upper respiratory tract infection. When triggered, the provider was presented with an alert offering the CDS tool upon opening the chart. If accepted, the provider was taken to a calculator with a list of clinical questions, each of which contributes to a total risk score (Figure 1). After calculator completion, the provider was shown a risk score, identifying the patient as low, intermediate, or high risk for the condition as well as offered an order set tailored to the calculated risk. These order sets included documentation for progress notes, laboratory orders, prescription orders, diagnoses, patient instructions, and level of service (Figure 2).

Live usability testing was conducted in a clinical office setting. Written informed consent was obtained from all participating providers the day before the study observations. At that time, the study procedures were reviewed with the providers and their staff. Testing was performed for 1 day for each of the providers. On the day of live usability testing, the providers’ receptionist handed out a flyer with details about the study to all of the

participating providers' patients. Study staff approached these patients to ask if they were being seen for a cough, sore throat, or an upper respiratory tract infection. Patients with these symptoms were provided with an explanation of the study and verbal consent was obtained.

All human-computer interactions, including audio and continuous screen capture, were recorded using Camtasia (TechSmith, Okemos, MI, USA) software. Before the start of the patient care session, the usability testing software was set to record. It was paused if patients left the room for testing and stopped at the end of the visit. After the provider's care sessions were completed, they were briefly interviewed about their general attitudes toward the tool. These interviews were recorded using a digital voice recorder.

All provider and patient verbalizations from the visits and the interviews were transcribed verbatim. The video from the visits, audio from the interviews, and the transcriptions of both underwent thematic analysis and were coded using the following process: a total of 2 coders used a triangulation approach involving iteratively watching the videos, listening to the interviews, and reading the transcriptions. This allowed a broader and more complex understanding of the data attained. Those 2 coders then undertook development of a codebook reflecting the emerging themes with no a priori codes used. Using the constant comparative method, additional readings of the transcription led to the consolidation of these coding schemes until no further refinement was required. The primary themes identified were: Tool Interruptions, Workflow, Tool Applicability, Patient-Tool interaction, Provider-Computer-Patient Interaction, Ease of Use, and Missed Opportunities. Transcribed audio from the visit and the interview along with observed participant interaction with the tool were coded by hand and were categorized under each code by 2 independent coders and analyzed for themes that would be generalizable to most CDS. The themes were reviewed together by the coders, and all discrepancies were resolved by discussion to achieve a consensus leading to 100% agreement between the coders. This was formative as opposed to summative usability testing. We did not measure task times, completion rates, or satisfaction scores. The institutional review board at the University of Wisconsin approved the research protocol.

Figure 1. Clinical decision support tool calculator.

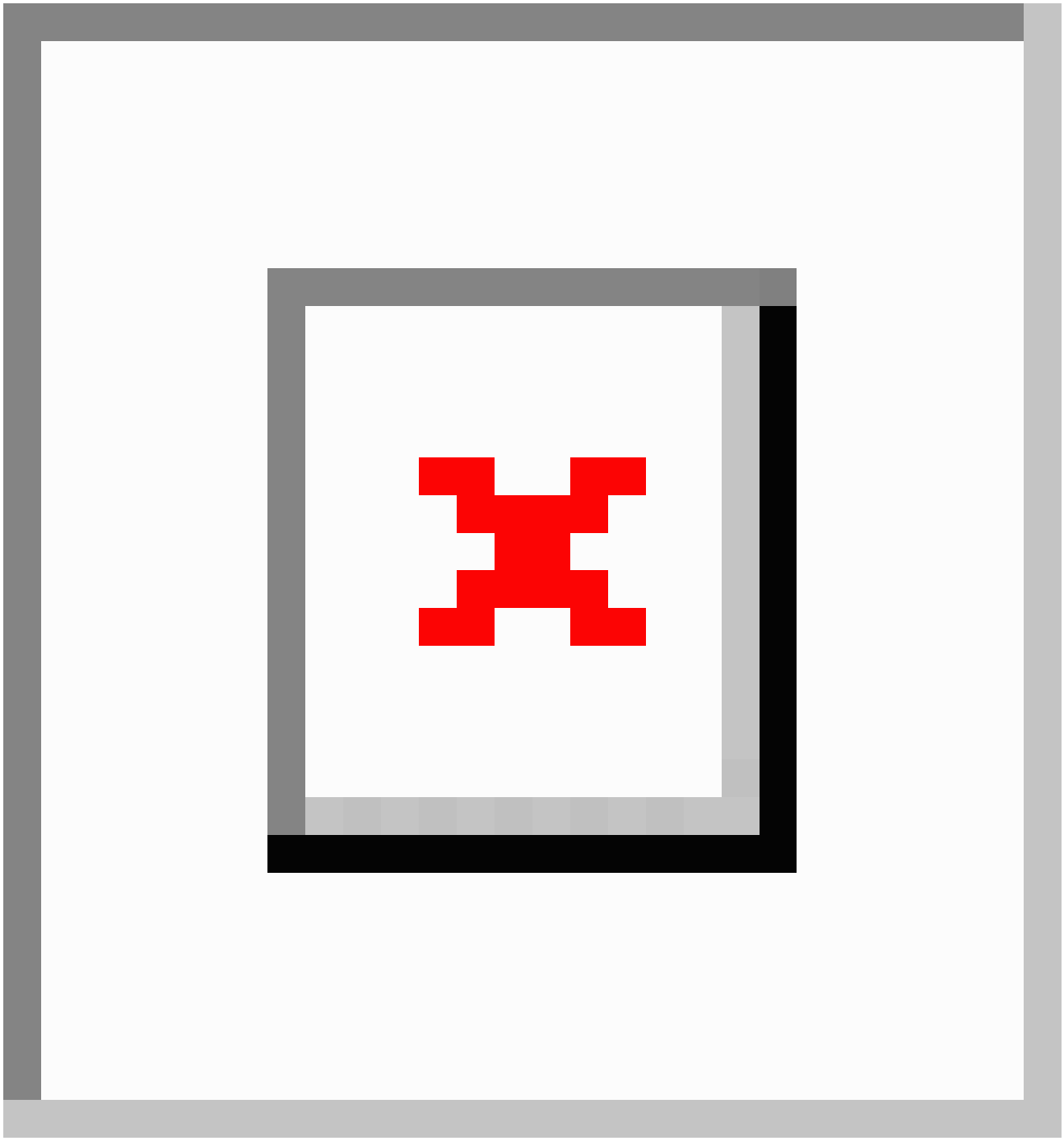
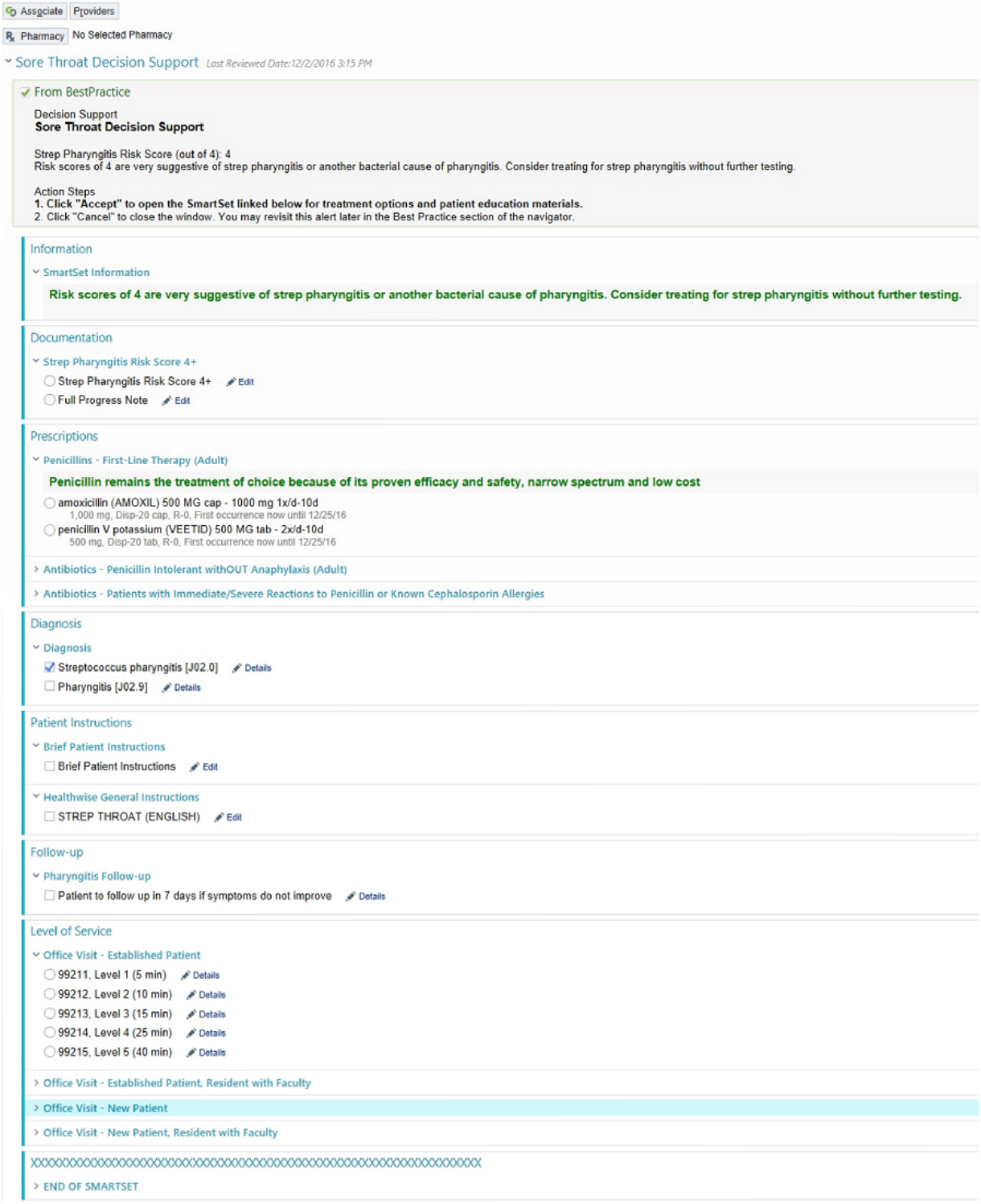


Figure 2. Clinical decision support tool automatic order set.



Results

Overview

All 3 participants were primary care providers: 2 nurse practitioners and 1 medical doctor. There were a total of 6 patient encounters. Although 5 of these were acute or follow-up visits that lasted about 15 min each, 1 was a complete physical exam that was about 30 min in length. In half of the visits, the patients presented with the chief complaint of sore throat, and the CDS tool built with the Centor Score was used to stratify the risk of GAS pharyngitis. In the other half of the visits, the patients presented with a chief complaint of cough or upper respiratory tract infection, and the CDS tool built with the Heckerling Rule was used to stratify the risk of pneumonia. As the tools were so similar, with the exception of clinical content, they were analyzed together. Example visit quotes, participant actions, and participant interview quotes are included in Table 1 by coding category along with a summary and recommendations for future CDS.

Table 1. Live usability testing results.

Coding category, example comments or actions[a]	Summary and recommendation
Tool interruptions	
Patient: "Was it last year or the year before – didn't I have to get a pneumonia shot?" *Provider navigates away from automatic order set immediately after opening it.*	During every testing session, the provider was interrupted during their use of the CDS[b] tool by the need to refer to other sections of the chart.
Provider: "Have you had a chest X-ray anytime recently?" *Provider clicks away from automatic order set to review results of last CXR*[c].	Recommendation: Complex CDS should be built for disrupted workflow, with easy and obvious re-entry points.
Workflow	
Provider opens chart, clicks away from alert, to progress notes.	During every testing session, the progress note served as the center point of the provider interaction with the electronic health record.
"It's the first thing that comes up...but you have to get all that info from the patient first. So that's what I mean by clunky." [PCI[d]]	—[e]
At the start of visit, all providers navigate immediately to the progress note. Half of them spent more than 95% of the visit with this function open, and only 1 spent more than 40% of the visit time with it open. [QM[f]]	Recommendation: CDS tools that exist within the progress note may have higher adoption rates because it would be more likely they were present at the time of decision making.
Tool applicability	
Provider: "So I read your chart; it says that you've been having symptoms as deer season?"	In half of the sessions, patient history challenged the validity of the clinical prediction rule used to calculate risk.
Patient: "I actually called in and Dr. [name] gave me a prescription..."	—
"Sometimes...something in your clinical encounter still says, 'get the X-ray or still treat,' you know, maybe you saw them before." [PCI]	Recommendation: CDS tools should be as broadly applicable as possible with clear indications for use.
Patient-tool interaction	
Provider: "OK, so our little risk calculator here is recommending that we would swab you for strep throat, and I agree with that."	In every session in which the tool was used to assess risk, the provider completed the calculator with the patient.
Provider: "But your heart is beating kinda fast, you've had a fever last night...the recommendation would be to get a chest x-ray today."	—
"I like to be able to show it to patients. So that part of it I really – I like to have that support, and that extra backup for the decision that I want to make." [PCI]	Recommendation: CDS tools should be designed to be viewed by the patient and provider simultaneously.
Provider-computer-patient interaction	
Patient: "My brother's living with me, he's a vet..." *Provider enters data from chart review into progress note while patient is talking about something unrelated.*	In every testing session, the providers toggled between addressing either the computer or the patient during the visit.
Provider: "So basically to summarize: about 9 days ago is when you first got sick..." *Physician stops interacting with computer to recap history.*	—
(Silence while physician types)	—
Providers spent 0% to 3% of their visit time listening to the patient without simultaneously engaging with the computer. [QM]	Recommendation: Providers may find CDS tools easier to complete if they engage patients.
Ease of use	
Provider: "Hold on, I just need the laboratory to actually put in the results... my thing isn't popping up for me to prescribe the antibiotics quite yet."	Providers were able to complete the tool quickly; however, during half of the sessions, hard stops and fixed elements in the tool created barriers to usability.
"The patient instructions have some hard stop, so I got frustrated with that, and then eventually deleted and typed my own patient instructions in." [PCI]	—
"Cause it's short. If it were any longer, I'd probably get frustrated with it." [PCI]	—
Providers spent about 1 min of the visit time completing the CDS tool. [QM]	Recommendation: Tools that are short, customizable, and flexible to different workflows will have improved usability.

Coding category, example comments or actions[a]	Summary and recommendation
Missed opportunities	
Provider enters shortcut ".cvuri" to generate upper respiratory infection note template at start of visit.	In every session, providers did not use either the automatic order set or automatic documentation.
Provider: "So the antibiotic that I would pick for you is one called Azithromycin." *Provider orders antibiotics a la carte without re-entering tool after chest x-ray is resulted.*	—
"It's easier for me to order a chest X-ray just outside of the order set...then get the results back and go on with the patient visit. And then at that point, it's like the opportunity has been lost to use the [automatic order] set." [PCI]	Recommendation: Elements that are incorporated into CDS tools as incentives should save the provider time or effort when compared with their usual workflow.

[a]Provider and patient statements during the visit are included in quotations, and provider actions are in italics.

[b]CDS: clinical decision support.

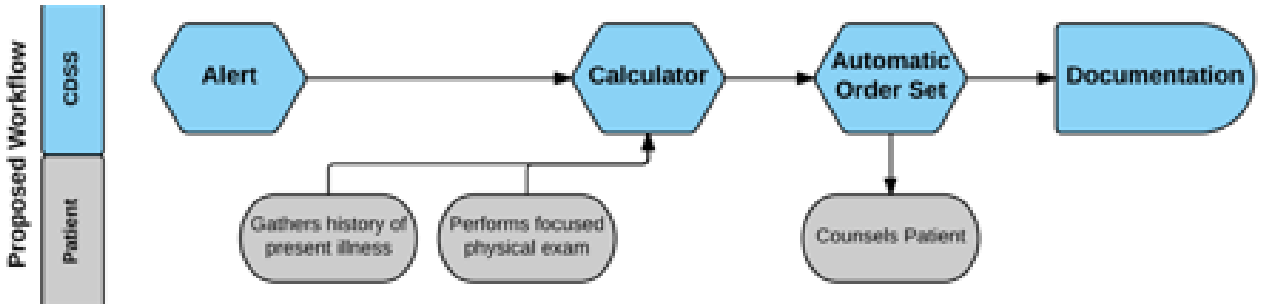
[c]CXR: chest x-ray.

[d]PCI: provider comments during interview.

[e]The Summary and Recommendation for each of the Coding Categories applies to all of the data provided.

[f]QM: quantitative measurements.

Figure 3. Clinical decision support system proposed workflow.



Coding Categories

Tool Interruptions

Although the tool was built to be completed sequentially and without interruption (Figure 3), all participants were interrupted during their use of the CDS tool. Participants were typically triggered to navigate away from the CDS tool by questions that came up during the encounter about the patient's previous medical history (eg, vaccine record and laboratory test results). Each of these deviations required the participant to remember to navigate back to the CDS tool and to know how to do this.

Workflow

Upon opening the chart, every participant was taken to an alert for the CDS tool. At the start of each patient session, the provider navigated away from the alert to the progress note and began taking the history of present illness. During most patient sessions, the provider then completed the physical exam, brought the patient back to the computer, and engaged with the CDS tool. The progress note served as the center point of the participant interaction with more than 95% of visit time spent with the progress note feature open in half of the sessions.

Tool Applicability

In half of the patient visits, patients reported some piece of information, typically as a part of the history of present illness that raised a question for the coders of whether the tool was applicable to their clinical condition. For example, 2 of the patient encounters were for complaints consistent with sinusitis and 1 patient with cough had been previously treated. All of the providers in the postsession brief interviews mentioned the value of a more broadly applicable tool that included CDS for bacterial sinusitis. They felt that this addition would allow them to use the tool more often.

Patient-Tool Interaction

A majority of the providers used the tool to assess risk by showing the patients the tool while they completed it and explained the results of the calculator to the patient. They all reported that the ability to show the patient their risk of a bacterial infection was the strongest feature of the tool. Providers reported using the tool to educate patients about their risk and manage patient expectations more than using it to discover the patient's risk of bacterial infection.

Provider-Computer-Patient Interaction

Providers spent most of the visit either talking to the patient or interacting with the EHR. They spent 0% to 3% of their time listening to the patient without engaging the EHR. For example, to gather the history of the present illness, providers typically started with an open-ended question. As the patient began talking, they shifted their focus to the EHR to begin typing the progress note. They took the opportunity to review the chart if the patient began talking about unrelated topics. At times when the patient was not speaking but the provider needed to interact

with the EHR (eg, completing orders at the end of the visit), there would be silence.

Ease of Use

Providers commented on the tool's brevity as being a significant strength, making it easier to use. They spent about 1 min of the patient visit completing the tool. Hard stops and fixed elements within the tool led to frustrations. For example, after a verbal communication about a positive rapid GAS pharyngitis result, the provider could not continue to the automatic order set until the result was properly registered by the laboratory, requiring the provider to leave the patient, go back to the laboratory, and resolve the issue before continuing with the patient visit.

Missed Opportunities

Although the tool was designed to automatically generate visit documentation as an incentive for tool completion, every provider started writing his or her note at the beginning of the visit. Each provider used shortcuts to template their notes, which increased the comparative ease of use of typing their note without using the tool's feature. Although the tool's automatic order set was also designed as an incentive for use, participants described it being easier to order antibiotics and tests outside of it.

Discussion

Principal Findings

This study contributes to our growing understanding of how to develop usable and useful CDS tools, particularly considering the provider-computer-patient interaction. This study builds on our previous work analyzing results from the "Think Aloud" and "Near Live" usability testing of these 2 CDS tools [38]. Each of these 3 types of usability testing generated unique and generalizable insights. As testing increasingly approached reality, additional types of barriers to and facilitators of CDS usage were found. During the "Think Aloud" testing, providers were presented with a written clinical case while interacting with the tool. Commentary focused on improving the ease of use of the tool. During the "Near Live" testing, providers interacted with a patient actor and commentary addressed ease of use of the tool with an added, more focused evaluation of its usefulness. Previous studies have also found that as usability testing approaches reality, themes and insights shift from mostly surface-level ease–of-use issues to high-level usefulness and workflow issues [28]. Live usability testing provided insights on the tools' ease of use, usefulness, and impact on the patient-provider interaction that were not evident in previous usability testing.

Provider-Computer-Patient Interaction and Patient-Tool Interaction

Our observation of the minimal time providers spent listening to the patient without simultaneously interacting with the computer speaks to the growing demands of the EHR. Each of these demands must take the place of some part of what was already a full visit. In a typical encounter, a provider listens to the patient, examines the patient, and talks to the patient. The pressure to "multitask" using the EHR is easiest while listening to the patient. Notably, however, there is evidence that providers are doing this without decreasing patient satisfaction or diminishing the patient-provider relationship [20]. The use of EHRs in the ambulatory setting also does not seem to decrease quality of care [39]. However, the EHR contains a wealth of information that has the potential to positively impact care. The simple, intuitive, and informational design of this tool allowed providers to use it with their patients, allowing the EHR to provide important information while reconnecting the patient and the provider.

CDS designers have largely focused on these tools' contribution to medical decision making without considering its collaborative nature. To varying degrees, every medical decision is a shared decision. CDS tools that are built to engage both patient and provider target both decision makers. Every provider in this study cited the ability to share the tool's results with the patient as its greatest strength. These providers did not need a better understanding of patient's risk of bacterial infection as much as they needed a better way to communicate this information to the patient. CDS that accounts for the patient's role in decision making may be used to facilitate shared decision making, which may improve usability, increase adoption rates, thereby resulting in improved quality of care.

Tool Interruptions, Usability, and Workflow

The expected workflow for the tool was not observed in any encounter, and the providers did not use the tool at the time it triggered. In addition, when the tool was used, they were unable to flow from alert to calculator to automatic order set as it was designed to be used. These findings point to the existence of significant provider workflow variability. Primary care provider workflow is not prespecified and emerges based on the unique interaction between the patient and the provider's agendas [40]. Our study points to a short, flexible, and customizable CDS tool as more usable. Locating the CDS inside the progress note may help to address tool interruptions and improve usability and workflow. The progress note seems to be the center point of provider interaction with the computer. For many providers, this would make the tool available at the time of decision making and present while they use the split screen to refer back to the chart when necessary.

Missed Opportunities and Tool Applicability

The ability to use the tool in as many clinical situations as possible increases its usefulness. Every provider commented on the utility of adding a tool addressing risk of bacterial sinusitis. This addition would allow providers to apply these tools to almost any symptoms of upper respiratory tract infection. The more broadly these tools apply, the more valuable they may be to providers. In half of the visits, patient history challenged the validity of the clinical prediction rule used to calculate the risk of bacterial infection. Usefulness was addressed as well with providers' lack of use of the incorporated incentives. Elements that are incorporated into CDS tools as incentives should save the provider time or effort when compared with their usual workflow. The lack of order set use can also limit the ability of the CDS to improve evidence-based patient care and influence the type of antibiotics ordered.

Usability testing of CDS helps to close the gap between its current and potential impact on providers, their interactions with patients, and the quality of care they give. Although the EHR's poor usability and interference with face-to-face patient care are prominent sources of professional dissatisfaction, providers still believe in the potential of this technology [23]. The concept of evidence-based clinical care revolutionized medicine by demanding that interventions be formally evaluated. We must evaluate CDS with this same rigorous approach; usability tested and refined CDS can address unforeseen consequences, decrease strain on the provider and the patient-provider interaction, and garner the adoption rates required to have a meaningful positive impact.

Limitations

As typical for usability studies, participants were a convenience sample of volunteers rather than a representative sample. They were identified based on their higher-than-average use of this CDS tool. This was done to ensure tool usage on the day of testing. These providers may have a more positive opinion of it or use it in a way that is fundamentally different from that of the average provider. Even in this subset of providers predisposed to high CDS use, the tool was not used as designed and created workflow frustration. These providers may also use the EHR more during patient encounters than average. The sample size for this study was small because of the inherent logistical difficulty of live usability testing in the real clinical environment. However, usability testing is typically performed in just 5 sessions as thematic saturation begins to occur at this point [35-37]. We reached thematic saturation during our study, observing consistent and recurring themes across all of our recorded sessions. During testing, participants were aware that they were being recorded and may have changed their behavior and reported observations because of being observed (the Hawthorne effect). This testing was done with just 1 EHR, EpicCare, which may limit generalizability. However, this is the most widely used EHR in the United States. All of these limitations are inherent to usability studies and represent standard practice.

Conclusions

Live usability testing of this CDS tool provided insights on its ease of use, usefulness, and its impact on the patient-provider interactions that were not evident in previous usability testing. This highlights the importance of incorporating live usability testing into CDS tool development. Our study suggests that short, flexible, and customizable CDS tools may be more usable, addressing the challenges of the highly variable provider workflow. The progress note seems to be the center point of provider interaction with the EHR. Locating the CDS tool inside the progress note may help to address tool interruptions and ensure that the tool is available at the time of decision- making and present when providers refer back to the chart when necessary. The tool was designed to be used sequentially and this contributed to providers not finishing the tool once they deviated from the intended workflow.

The more broadly these tools apply, the more valuable they are to providers. Elements that are incorporated into CDS tools as incentives must be useful, saving the provider time or effort when compared with their usual workflow. Live usability testing of these tools also generated insights about their impact on the patient-provider interaction. The simple, intuitive, and informational design of the tool allowed providers to use it with their patients. CDS can contribute to the patient-provider interaction by being built to be simultaneously viewed by the provider and patient. The use of the calculator to engage the patient in the decision-making process as a driver for the use of the CDS tool needs further study. This allows the EHR to provide important information while reconnecting patient and provider.

Acknowledgments

This project was funded by the National Institutes of Health, National institute of Allergy and Infectious Diseases, under grant #5R01 AI108680-03 and the National Heart Lung and Blood Institute under grant # 1K23HL145114-01. The funding body had no role in the design of the study or the collection, analysis, or interpretation of data.

Conflicts of Interest

None declared.

References

1. Kohn LT, Cor JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000.
2. Connelly DP, Rich EC, Curley SP, Kelly JT. Knowledge resource preferences of family physicians. J Fam Pract 1990 Mar;30(3):353-359. [Medline: 2248632]
3. Institute of Medicine. Crossing The Quality Chasm: A New Health System For The 21st Century. Washington, DC: National Academy Press; 2001.
4. Asch S, McGlynn EA, Hogan MM, Hayward RA, Shekelle P, Rubenstein L, et al. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. Ann Intern Med 2004 Dec 21;141(12):938-945. [doi: 10.7326/0003-4819-141-12-200412210-00010] [Medline: 15611491]
5. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]

6. Adler-Milstein J, Jha AK. HITECH Act drove large gains in hospital electronic health record adoption. Health Aff (Millwood) 2017 Dec 1;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
7. Bernstein S, Whitaker D, Winograd J, Brennan JA. An electronic chart prompt to decrease proprietary antibiotic prescription to self-pay patients. Acad Emerg Med 2005 Mar;12(3):225-231 [FREE Full text] [doi: 10.1197/j.aem.2004.09.021] [Medline: 15741585]
8. Garthwaite E, Will EJ, Bartlett C, Richardson D, Newstead CG. Patient-specific prompts in the cholesterol management of renal transplant outpatients: results and analysis of underperformance. Transplantation 2004 Oct 15;78(7):1042-1047. [doi: 10.1097/01.TP.0000137340.22880.C8] [Medline: 15480172]
9. Gaikwad R, Sketris I, Shepherd M, Duffy J. Evaluation of accuracy of drug interaction alerts triggered by two electronic medical record systems in primary healthcare. Health Informatics J 2007 Sep;13(3):163-177. [doi: 10.1177/1460458207079836] [Medline: 17711879]
10. Seidling HM, Schmitt SPW, Bruckner T, Kaltschmidt J, Pruszydlo MG, Senger C, et al. Patient-specific electronic decision support reduces prescription of excessive doses. Qual Saf Health Care 2010 Oct;19(5):e15. [doi: 10.1136/qshc.2009.033175] [Medline: 20427312]
11. Shah NR, Seger AC, Seger DL, Fiskio JM, Kuperman GJ, Blumenfeld B, et al. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc 2006;13(1):5-11 [FREE Full text] [doi: 10.1197/jamia.M1868] [Medline: 16221941]
12. Smith D, Perrin N, Feldstein A, Yang X, Kuang D, Simon SR, et al. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. Arch Intern Med 2006 May 22;166(10):1098-1104. [doi: 10.1001/archinte.166.10.1098] [Medline: 16717172]
13. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J, et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. CMAJ 2003 Sep 16;169(6):549-556 [FREE Full text] [Medline: 12975221]
14. Bright T, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 3;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
15. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Arch Intern Med 2003 Jun 23;163(12):1409-1416. [doi: 10.1001/archinte.163.12.1409] [Medline: 12824090]
16. Bonnabry P, Despont-Gros C, Grauser D, Casez P, Despond M, Pugin D, et al. A risk analysis method to evaluate the impact of a computerized provider order entry system on patient safety. J Am Med Inform Assoc 2008;15(4):453-460 [FREE Full text] [doi: 10.1197/jamia.M2677] [Medline: 18436900]
17. Roshanov P, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. Br Med J 2013 Feb 14;346:f657 [FREE Full text] [doi: 10.1136/bmj.f657] [Medline: 23412440]
18. Souza N, Sebaldt RJ, Mackay JA, Prorok JC, Weise-Kelly L, Navarro T, CCDSS Systematic Review Team. Computerized clinical decision support systems for primary preventive care: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011 Aug 3;6:87 [FREE Full text] [doi: 10.1186/1748-5908-6-87] [Medline: 21824381]
19. McGinn T, McCullagh L, Kannry J, Knaus M, Sofianou A, Wisnivesky JP, et al. Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. JAMA Intern Med 2013 Sep 23;173(17):1584-1591. [doi: 10.1001/jamainternmed.2013.8980] [Medline: 23896675]
20. 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. [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
21. Murphy D, Meyer AN, Russo E, Sittig DF, Wei L, Singh H. The burden of inbox notifications in commercial electronic health records. JAMA Intern Med 2016 Apr;176(4):559-560 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0209] [Medline: 26974737]
22. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
23. Friedberg M, Chen PG, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. Rand Health Q 2014;3(4):1 [FREE Full text] [Medline: 28083306]
24. Bates D, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530 [FREE Full text] [doi: 10.1197/jamia.M1370] [Medline: 12925543]

25. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson SA. Computerized clinical decision support for prescribing: provision does not guarantee uptake. J Am Med Inform Assoc 2010;17(1):25-33 [FREE Full text] [doi: 10.1197/jamia.M3170] [Medline: 20064798]
26. Graham T, Kushniruk AW, Bullard MJ, Holroyd BR, Meurer DP, Rowe BH. How usability of a web-based clinical decision support system has the potential to contribute to adverse medical events. AMIA Annu Symp Proc 2008 Nov 6:257-261 [FREE Full text] [Medline: 18998968]
27. Payne T, Hines LE, Chan RC, Hartman S, Kapusnik-Uner J, Russ AL, et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. J Am Med Inform Assoc 2015 Nov;22(6):1243-1250. [doi: 10.1093/jamia/ocv011] [Medline: 25829460]
28. Li A, 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]
29. Devine E, Lee CJ, Overby CL, Abernethy N, McCune J, Smith JW, et al. Usability evaluation of pharmacogenomics clinical decision support aids and clinical knowledge resources in a computerized provider order entry system: a mixed methods approach. Int J Med Inform 2014 Jul;83(7):473-483 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.04.008] [Medline: 24874987]
30. Kastner M, Lottridge D, Marquez C, Newton D, Straus SE. Usability evaluation of a clinical decision support tool for osteoporosis disease management. Implement Sci 2010 Dec 10;5:96 [FREE Full text] [doi: 10.1186/1748-5908-5-96] [Medline: 21143978]
31. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. J Biomed Inform 2012 Dec;45(6):1202-1216 [FREE Full text] [doi: 10.1016/j.jbi.2012.09.002] [Medline: 22995208]
32. Fossum M, Ehnfors M, Fruhling A, Ehrenberg A. An evaluation of the usability of a computerized decision support system for nursing homes. Appl Clin Inform 2011;2(4):420-436 [FREE Full text] [doi: 10.4338/ACI-2011-07-RA-0043] [Medline: 23616886]
33. Kortteisto T, Komulainen J, Mäkelä M, Kunnamo I, Kaila M. Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care. BMC Health Serv Res 2012 Oct 8;12(1):349 [FREE Full text] [doi: 10.1186/1472-6963-12-349] [Medline: 23039113]
34. Feldstein D, Hess R, McGinn T, Mishuris RG, McCullagh L, Smith PD, et al. Design and implementation of electronic health record integrated clinical prediction rules (iCPR): a randomized trial in diverse primary care settings. Implement Sci 2017 Dec 14;12(1):37 [FREE Full text] [doi: 10.1186/s13012-017-0567-y] [Medline: 28292304]
35. 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]
36. Virzi R. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 2016 Nov 23;34(4):457-468. [doi: 10.1177/001872089203400407]
37. Nielsen J. Nielsen Norman Group. 2012. How many test users in a usability study URL: https://www.nngroup.com/articles/how-many-test-users/ [accessed 2019-03-18] [WebCite Cache ID 76xiLhzvO]
38. Richardson S, Mishuris R, O'Connell A, Feldstein D, Hess R, Smith P, et al. "Think aloud" and "Near live" usability testing of two complex clinical decision support tools. Int J Med Inform 2017 Dec;106:1-8 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.06.003] [Medline: 28870378]
39. Linder J, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. Arch Intern Med 2007 Jul 9;167(13):1400-1405. [doi: 10.1001/archinte.167.13.1400] [Medline: 17620534]
40. Holman G, Beasley JW, Karsh BT, Stone JA, Smith PD, Wetterneck TB. The myth of standardized workflow in primary care. J Am Med Inform Assoc 2016 Jan;23(1):29-37 [FREE Full text] [doi: 10.1093/jamia/ocv107] [Medline: 26335987]

Abbreviations

CDS: clinical decision support
EHR: electronic health record
GAS: group A Streptococcus
HITECH: Health Information Technology for Economic and Clinical Health Act

Edited by G Eysenbach; submitted 16.10.18; peer-reviewed by M Horrocks, R Alturki; comments to author 06.12.18; revised version received 10.01.19; accepted 30.01.19; published 15.04.19.

Please cite as:
Richardson S, Feldstein D, McGinn T, Park LS, Khan S, Hess R, Smith PD, Mishuris RG, McCullagh L, Mann D
Live Usability Testing of Two Complex Clinical Decision Support Tools: Observational Study
JMIR Hum Factors 2019;6(2):e12471
URL: http://humanfactors.jmir.org/2019/2/e12471/
doi:10.2196/12471
PMID:30985283

Original Paper

Interruptive Versus Noninterruptive Clinical Decision Support: Usability Study

Saul Blecker[1,2], MD, MHS; Rishi Pandya[2], MD; Susan Stork[1]; Devin Mann[1], MD, MPH; Gilad Kuperman[3], MD, PhD; Donna Shelley[1], MD, MPH; Jonathan S Austrian[2], MD

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

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

[3]Memorial Sloane Kettering Cancer Center, New York, NY, United States

Corresponding Author:
Saul Blecker, MD, MHS
Department of Population Health
New York University School of Medicine
227 E 30th St, 6th Floor
New York, NY, 10016
United States
Phone: 1 6465012513
Email: saul.blecker@nyumc.org

Abstract

Background: Clinical decision support (CDS) has been shown to improve compliance with evidence-based care, but its impact is often diminished because of issues such as poor usability, insufficient integration into workflow, and alert fatigue. Noninterruptive CDS may be less subject to alert fatigue, but there has been little assessment of its usability.

Objective: This study aimed to study the usability of interruptive and noninterruptive versions of a CDS.

Methods: We conducted a usability study of a CDS tool that recommended prescribing an angiotensin-converting enzyme inhibitor for inpatients with heart failure. We developed 2 versions of the CDS: an interruptive alert triggered at order entry and a noninterruptive alert listed in the sidebar of the electronic health record screen. Inpatient providers were recruited and randomly assigned to use the interruptive alert followed by the noninterruptive alert or vice versa in a laboratory setting. We asked providers to "think aloud" while using the CDS and then conducted a brief semistructured interview about usability. We used a constant comparative analysis informed by the CDS Five Rights framework to analyze usability testing.

Results: A total of 12 providers participated in usability testing. Providers noted that the interruptive alert was readily noticed but generally impeded workflow. The noninterruptive alert was felt to be less annoying but had lower visibility, which might reduce engagement. Provider role seemed to influence preferences; for instance, some providers who had more global responsibility for patients seemed to prefer the noninterruptive alert, whereas more task-oriented providers generally preferred the interruptive alert.

Conclusions: Providers expressed trade-offs between impeding workflow and improving visibility with interruptive and noninterruptive versions of a CDS. In addition, 2 potential approaches to effective CDS may include targeting alerts by provider role or supplementing a noninterruptive alert with an occasional, well-timed interruptive alert.

(JMIR Hum Factors 2019;6(2):e12469) doi:10.2196/12469

KEYWORDS

clinical decision support; hospital; electronic health records

Introduction

Clinical decision support (CDS) systems have been shown to improve provider compliance with evidence-based cardiovascular care in the inpatient hospital setting [1,2]. Nonetheless, the effectiveness of CDS interventions is frequently diminished because of issues such as poor usability, insufficient integration into provider workflow, and alert fatigue [3-6]. These limitations are particularly problematic in the inpatient setting, where providers are concurrently caring for numerous patients with urgent needs located in multiple locations.

CDS alert fatigue is frequently related to the fact that interruptive alerts force providers to notice or respond to the CDS in the middle of their other tasks. Noninterruptive CDS tools, which do not require stoppage of other electronic health record (EHR) activity, may be less subject to alert fatigue [6,7]. A number of studies have demonstrated that noninterruptive alerts can increase provider compliance with care measures such as venous thromboembolism prevention in the inpatient setting [8,9], yet this type of alert is generally perceived as less successful at changing provider behavior compared with interruptive alerts [6,7]. Nonetheless, the few studies that compare the relative uptake of interruptive and noninterruptive alerts have not consistently shown interruptive alerts to be superior [10,11]. Finally, despite assumptions that noninterruptive alerts have less effect on workflow [6,7], there has been little evaluation of the relative usability of interruptive and noninterruptive alerts. Although prior studies have evaluated the usability of either interruptive or noninterruptive alerts [9,12,13], we are unaware of studies that have compared the usability of these 2 implementation approaches. Information about their relative usability can help inform developers of CDS about the relative advantages of interruptive and noninterruptive alerts. In addition, evaluation of usability of these 2 CDS implementation approaches may be particularly useful in the inpatient setting, where providers frequently deal with competing demands and interruptions to workflow.

Usability relates to the extent a system will allow end users to complete a task in an effective, timely, and satisfactory way [14,15]. Usability testing draws on the principles of human-computer interaction to evaluate the usability of a system and is considered best practice in the development of EHRs and related systems in health care [14-17]. The purpose of this study was to pilot test the comparative usability of an interruptive version versus a noninterruptive version of an inpatient-focused CDS.

Methods

Study Design

We conducted a usability study of a CDS tool that recommended prescribing an angiotensin-converting enzyme (ACE) inhibitor for inpatients with heart failure. The setting was NYU Langone Health, an urban academic medical center with approximately 3000 hospitalizations with a diagnosis of heart failure annually [18]. We recruited individual health care providers to use the tool in a laboratory setting and provide feedback on usability. We created 2 versions of the CDS: one an interruptive alert and the other a noninterruptive alert. We then randomly assigned providers to use the interruptive alert followed by the noninterruptive alert or vice versa; we randomly assigned the order for presentation to minimize the effect that using one version of the alert may have on feedback on the second version of the alert. Order assignment was based on random number generation. Following usability assessment, we conducted a brief semistructured interview for additional feedback.

Subjects and Recruitment

We included individual health care providers who care for and write medication orders for hospitalized adult patients. We excluded providers who do not write inpatient medication orders. We identified and recruited potential participants through sending emails to relevant department listservs, colleagues of study team members, and suggestions from prior interviewees. We used a purposive sampling framework: we invited participants to ensure a range of services, including medicine and surgery, and provider types, including attending physicians, resident physicians, nurse practitioners (NPs), and physician assistants. However, we stopped recruiting attending physicians after the first interview, in which the attending physician reported exclusively relying on residents, NPs, and physician assistants to write orders. Recruitment continued until a range of services and provider types was achieved and thematic saturation was reached. Participants received a US $25 gift card after completion of the interview.

Clinical Decision Support Intervention Description

We developed 2 versions of the CDS intervention that had similar triggering actions but varied in their format for presentation. The CDS interventions were built within the sandbox testing environment of the EHR at NYU Langone Health, Epic (Epic Systems, Verona, Wisconsin). The initial development was led by the study team using input from clinical leadership and based on standard Epic templates. Development was informed by interviews with end-user providers [19]. Both versions of the CDS ultimately presented a dialogue box that informed providers that the patient had a reduced ejection fraction (EF), was not on an ACE inhibitor or angiotensin receptor blocker, and that these medications are potentially beneficial to patients with this condition [20,21]. Usual contraindications were explained, and recent values for blood pressure, estimated glomerular filtration rate (eGFR), and potassium were listed [20]. Providers were given the options to order an ACE inhibitor (lisinopril 5 mg daily), report a contraindication, or simply dismiss the CDS.

The format of the first alert was interruptive, in which the CDS dialogue box popped up at the time of order entry (Figure 1). The second version was a noninterruptive link that was located in a sidebar *checklist report* (Figure 2). This sidebar was part of the usual EHR display, and the interruptive alert was present in the sidebar until the CDS criteria were satisfied. Selecting the hyperlink in the sidebar led to the presentation of the same CDS dialogue box as in the interruptive alert.

Usability Testing

We first obtained verbal consent for participation and audio recording. We then provided participants with a clinical scenario in which they were caring for a patient who had heart failure with a reduced EF and who was principally hospitalized for another condition related to the provider's specialty, such as pneumonia, stroke, or surgery. Providers were advised that they were to proceed with opening the patient's chart and ordering morning laboratory tests. For providers assigned to the interruptive alert, the alert would trigger once they initiated the process to order labs. Some providers who were first assigned to the noninterruptive alerts would see and attempt to work with the CDS as well; for those who had not noticed the noninterruptive alert after a few minutes of charting, we also directed them to the CDS. While working with the CDS,

providers were asked to *think aloud* [17,22]. In the think-aloud method, users verbalize their thoughts and offer feedback while interacting with the CDS to identify usability issues.

After working through the first alert, we asked providers about navigation, content, ease of use, fit into workflow, and suggestions. We then performed usability testing on the other version of the alert using the same procedure. Providers were then asked about usability of the second version of the alert and the comparative advantages of each version of the CDS tool. Finally, we asked providers to complete a brief demographic survey.

Qualitative Analysis

Audio recordings from usability testing were transcribed by a professional transcription service. Transcriptions were reviewed against recordings, with any mistakes corrected.

Figure 1. Screenshot of clinical decision support used in usability testing: interruptive version of clinical decision support. Source: Epic Systems Corporation; used with permission. ACE: angiotensin-converting enzyme; ARB: angiotensin receptor blocker; BP: blood pressure; eGFR: estimated glomerular filtration rate; LVEF: left ventricular ejection fraction.

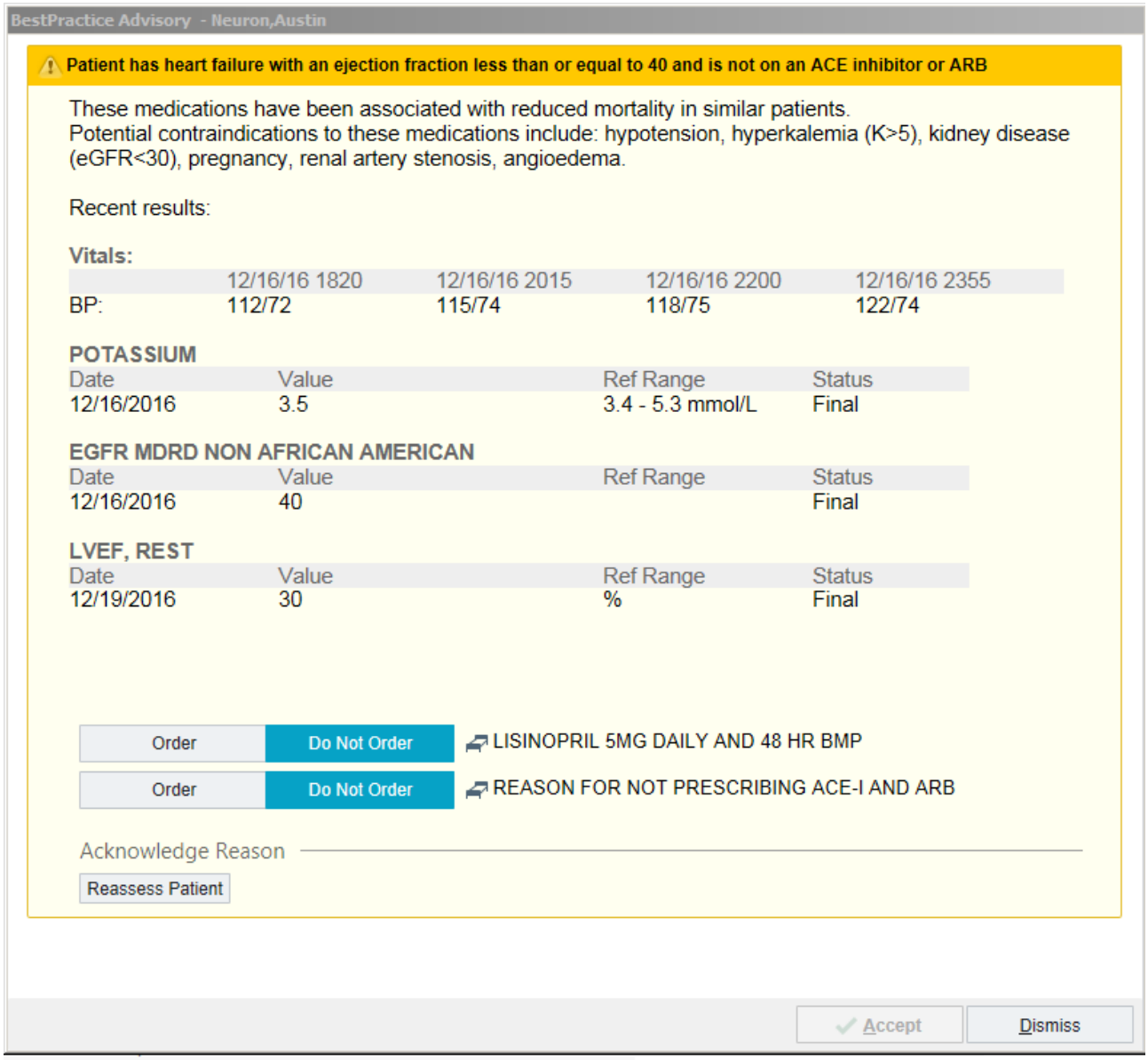
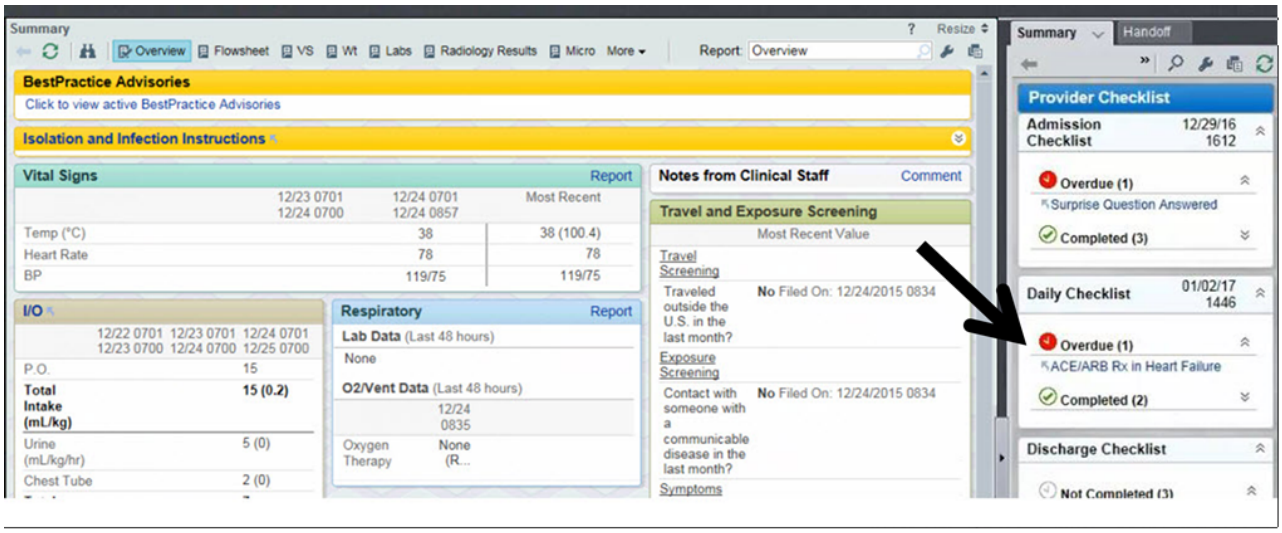


Figure 2. Screenshot of clinical decision support used in usability testing: location of noninterruptive version of clinical decision support, highlighted by the arrow. Clicking the link in the noninterruptive alert would take the user to a screen similar to the interruptive alert. Source: Epic Systems Corporation; used with permission. ACE: angiotensin-converting enzyme; ARB: angiotensin receptor blocker; BP: blood pressure.



We used the constant comparative method to analyze the usability testing, which included audio recordings from both the think-aloud protocol and the semistructured interview questions. In this qualitative analysis technique based on grounded theory [23,24], we began with open codes that were progressively grouped and refined into code categories. At least 2 of 3 coders (SB, RP, and SS) independently coded each of the transcripts and then met to review codes and reach consensus on any disagreements. Throughout the coding process, coders also met regularly with the larger investigative team to review and refine the code list. We categorized codes as being related to the general CDS, the interruptive alert, and the noninterruptive alert. Emergent themes were informed by the CDS Five Rights CDS framework as well as by prior work in CDS usability testing [13,15,25]; the Five Rights framework postulates that CDS is most effective when the right information is delivered to the right person, through the right intervention format and the right channel, and at the right time in workflow. We grouped all codes into 1 of the 4 rights in the framework; no codes were related to the theme of right channel as the CDS was delivered exclusively through the EHR.

Results

We conducted usability testing with 12 providers. Overall, 9 of these providers (75%) were on the medicine service; the remaining providers were in surgery or neurology (Table 1). Half of the providers self-identified as Asian. Furthermore, 7 providers were randomly assigned to test the usability of the interruptive alert followed by the noninterruptive alert; the remaining 5 providers started with the noninterruptive alert. Interview lengths ranged from 11 to 29 min.

We categorized codes from the usability testing, which combined the think-aloud interviews and the responses to semistructured questions, into 4 themes related to the CDS Five Rights. We defined some codes as related to the CDS in general and others as related to the interruptive or noninterruptive version of the CDS (Textbox 1 and Table 2).

Right Information

Nearly all codes that related to the general CDS fit within the theme of right information and could generally be categorized as positives, negatives, or suggestions for the CDS. Positives about the CDS included that providers thought that the CDS would be helpful to alert them that the patient had heart failure and may not be on evidence-based therapy. They expressed that the reported contraindications in the text and reporting of the relevant vital signs and laboratory results were useful. Nonetheless, providers suggested adding the following elements: trends for laboratory results, creatinine results, summary of past ACE inhibitor use, and contraindications to ACE inhibitors. Concurrently, some providers gave negative feedback about too much information, which could impede workflow, as suggested by 1 first-year resident:

> *I don't know if there's a way to make it even more brief...there's too much text...it was slowing me down.*

The primary negative feedback related to usability stemmed from a lack of clarity on the slide button that allowed for options to *order* or *do not order* each of the ACE inhibitor or reason for not prescribing. A second-year resident made this suggestion:

> *It was a little confusing...I don't know if there's a way to make it to that so it's order or not order the ACE, and then the second one [for] if you didn't order the ACE.*

Table 1. Characteristics of 12 providers participating in usability testing.

Characteristic	Statistics, n (%)
Clinical service	
Medicine	9 (75)
Neurology	1 (8)
Surgery	2 (17)
Clinical role	
Attending	1 (8)
Resident	—[a]
First-year resident	2 (17)
Second-year resident	4 (33)
Nurse practitioner	4 (33)
Physician assistant	1 (8)
Years in current role	
1-3	6 (50)
4-10	4 (33)
>10	2 (17)
Female	4 (33)
Ethnicity	
Not Hispanic or Latino	11 (92)
Missing	1 (8)
Race	
White	3 (25)
Black	1 (8)
Asian	6 (50)
Multiracial	2 (17)

[a]Not applicable.

Right Person

Perceived role, a contributor to the theme of right person, also influenced whether providers found the general CDS tool to be useful. In particular, a number of providers felt that it was their responsibility to deliver evidence-based care, including for an ACE inhibitor in heart failure. However, some providers, including those on surgical services and those who perform cross-coverage duties, found the CDS to be outside of their scope of practice. These providers wished for the option to dismiss the CDS for themselves but not for other providers. In this approach, the CDS would only continue triggering providers whose perceived role was appropriate for the CDS recommendations.

The theme of right person also applied to each version of the alert. A second-year medicine resident preferred the noninterruptive alert because of their perceived role to conduct global reviews of their patients:

> *At the end of the day I look through every [patient's checklist] as a [senior] resident. As [a first year resident] maybe not because I'm the one putting in all the orders.*

A first-year resident felt the interruptive alert would be useful because:

> *In the acute setting, especially, Lisinopril might get missed until we discuss it during rounds, but then if you put that there as an alert for us to see. [If I am too busy to] order at that time I feel like I would write it down somewhere to keep myself...I keep a sheet with all the patients and to-dos for every patient...and I definitely won't forget that because I know by the end of the day I want to check off all the boxes.*

Textbox 1. Codes from usability testing categorized into themes based on the Clinical Decision Support Five Rights. Groups further based on the interruptive or noninterruptive version of alert or usability groups.

Right information

- Content/usefulness
  - Alerts to best practice
  - Alerts to presence of heart failure
  - Labs and vital signs relevant
  - Contraindications to therapy useful
  - Wants a summary of current and prior medications
  - Wants lab or vital sign trends
  - No creatinine listed
  - Less information would be helpful
- Usability
  - Easy to locate relevant information
  - Difficulty or confusion with “order” versus “do not order” button
  - Does not notice the reason for not prescribing
  - Difficulty with ordering basic labs within clinical decision support

Right person

- General alert
  - Recommendation not within the perceived scope of practice
  - Responsibility to deliver evidence-based therapy
- Noninterruptive version of alert
  - Responsibility as a resident to address noninterruptive alerts

Right time in workflow

- Noninterruptive version of alert
  - Likes ability to address at a later time
  - Reviewing noninterruptive alerts part of workflow
- Interruptive version of alert
  - Prefers if delivered at right time in workflow
  - Impedes workflow
  - Wants option to address at a later time

Right intervention format

- Noninterruptive version of alert
  - Not always noticed
  - Flagged alerts increase visibility
  - Likes that can defer task to another provider
  - May defer and then forget about alert
  - Would notice and address noninterruptive alert
  - Would prefer alerts in a more visible location
- Interruptive version of alert

- More noticeable
- Pays less attention to content of interruptive alert

- Either version
  - Prefers combination of interruptive and noninterruptive alert

Table 2. Example quotations from usability testing by an interruptive or noninterruptive version of the clinical decision support.

Theme and clinical decision support version	Example code	Example quotation
Right information		
General (content)	Wants lab or vital sign trend	"It's better to have a trend...I'm more comfortable ordering this because I see those three times patient is very stable"
General (usability)	Difficulty with "order" versus "do not order" button	"To me that's a little counter-intuitive, but it could be that there's other sections of [the EHR] where that's how you document not doing something."
Right person		
General	Not within the perceived scope of practice	"I wouldn't necessarily start a patient on a medication just because of my specialty."
Noninterruptive	Responsibility as a resident to address noninterruptive alerts	"I started using the provider checklist a little bit more especially as a resident when you're reviewing things."
Right time in workflow		
Noninterruptive	Reviewing alert part of workflow	"At the end of the day I look through everyone, make sure...things are checked. Then I would notice things that are here."
Interruptive	Impedes workflow	"This one is a little bit more annoying because it will prevent me from doing what I wanna do."
Right intervention format		
Noninterruptive	Not always noticed	"If you hadn't have told me that this was on the right-hand side, I never would have noticed it in the first place. Now that I see it here it's actually nice."
Interruptive	Pays less attention to content of interruptive	"When we get a lot of them we tend to just turn off—when I see it I just barely breeze right through it and just hit dismiss."
Either version	Combine interruptive and noninterruptive	"I don't know if there's any way to make it pop up if you haven't reviewed the provider checklist by the end of the day."

Right Time in Workflow

A number of providers expressed that they preferred the noninterruptive version as it fit better within their workflow by not impeding current tasks. They expressed appreciation that they could address a noninterruptive alert at a later point, according to their own workflow. Some providers expressed that reviewing noninterruptive CDS tools was part of their current daily routine. Conversely, many providers agreed with an NP who described the interruptive CDS as *annoying* and impeded workflow. They requested the capability to defer the alert until a later time but did note that the interruptive version would be preferred if activated at the right time in their workflow.

Right Intervention Format

Many providers said that they do not always notice noninterruptive alerts or that they defer these alerts and forget to return to them at a later time. Others found the location on the screen to have inadequate visibility or believed that there were too many flagged alerts on the screen, making the alert less noticeable. Indeed, over half of the providers (7/12) did not quickly notice the noninterruptive alert and were directed to its location on the screen; of these providers, 3 were initially assigned to the noninterruptive alert and 4 were initially assigned to the interruptive alert. Conversely, some providers found the noninterruptive alert flags to be readily visible and appreciated the ability to defer these alerts to another time and, if appropriate, to another provider. Providers found the interruptive alert to be more noticeable, which is why some preferred this format; concurrently, others said they would not pay attention to the interruptive alert, including 1 second-year resident:

> *I feel making it mandatory makes it like I'd pay attention to it less.*

Given the noted trade-offs between the 2 versions of these alerts, some providers thought a combined version would be most useful. For instance, the noninterruptive alert could be available continuously but, if not utilized within a certain timeframe,

would be enhanced with triggering of the interruptive alert, as described by 1 second-year resident:

> *You have to electively review the [noninterruptive alert], which everyone may or may not do...I don't know if there's any way to make it pop up if you haven't reviewed by the end of the day or [another] time frame.*

Discussion

Principal Findings

We found that many providers expressed annoyance in working with an interruptive CDS, primarily because it would interrupt workflow. A noninterruptive version of the CDS was appealing to providers, given that it could be accessed at any time in the workflow or seamlessly deferred to other providers. However, providers acknowledged that a noninterruptive alert was frequently not noticed and may not support clinical decision making unless integrated into routine workflow. One suggestion was to balance the 2 approaches by combining formats: supplementing a noninterruptive alert with an occasional, well-timed interruptive alert if uptake was insufficient. Given the reported trade-off of distraction and visibility between the interruptive and noninterruptive alerts, we intend on implementing both versions of the CDS in our hospital system to determine relative use and usability in clinical practice.

Although individual providers differed on their description of how each version of the alert would fit into their workflow, one of our key findings was that provider role seemed to be associated with the acceptability of the CDS format. In particular, some providers expressed that their role in residency training affected their preference for how the CDS was delivered. With the caveat that this small qualitative study was not powered to represent subgroups, we found that 1 first-year resident, whose role is primarily related to implementation of the care that is delivered in the hospital, tended to favor the interruptive CDS as it alerted this provider to another task to accomplish for the day. Conversely, more senior residents, whose role is defined by overseeing the delivery of care for patients, tended to favor the noninterruptive CDS. These residents felt that such a CDS could aid in their broad assessment of an individual patient's care at the opportune time when performing such a review.

Our finding of a potential interaction between provider role and fit of CDS into workflow builds off prior studies examining provider characteristics and potential for uptake of CDS [26-28]. For instance, surveyed providers were more likely to report acceptance of a CDS if not behind in their work [28], and in secondary analysis, a CDS tool for respiratory symptoms was more likely to be used by resident providers as compared with attending providers [27]. The CDS Five Rights framework specifies the importance of both provider role and intervention format [25]. This framework has led to CDS systems designed to deliver different information for clinicians in different roles; for example, 1 CDS system included an alert to nurses if a patient had signs of early sepsis while concurrently offering a separate sepsis order set to providers [29]. Nonetheless, we are unaware of a CDS that was developed to specifically address the potential interaction between role and intervention format. Our data suggest an opportunity to increase CDS usability—ultimately with the goal of increasing uptake—by targeting providers who may find that a given format fits best within their clinical role. An example of this based on our preliminary findings could be that the interruptive alert targets first-year residents, whereas the noninterruptive alert targets senior residents; however, we would need to better survey residents before such an implementation.

One of the primary purposes of usability testing of a CDS tool is to adjust the tool based on end-user feedback [15]. We made some changes to our CDS during usability testing based on initial feedback, including incorporating additional trends in blood pressure and laboratory results. We only later made the suggested change of adding creatinine to the CDS. We were initially resistant to changing this laboratory presentation, as guidelines recommend eGFR—rather than creatinine—as the preferred method for evaluating kidney function [30]. We eventually added creatinine, given the consistent request by end users. Further assessment of usability and uptake of eGFR in practice is warranted. There were also some suggestions that did not result in changes to the CDS. Although 1 suggestion was to list patient medications, we did not choose to do this because of concern related to a conflicting code of *too much information*. We also encountered some suggestions for which we had difficulties with changing the CDS. The biggest limitation in usability was the difficulty with using the order button; problems with this button resulted in some providers not ordering an ACE inhibitor even when they had intended to do so. However, this button was part of the native functionality of the vendor's EHR alert, and we were advised by our information technology department that the display of the order button was not configurable.

Limitations

A number of limitations should be considered in the interpretation of the results. First, the study was based in 1 institution and using a single EHR, so results may not be generalizable. Second, usability testing took place in a laboratory setting rather than in the context of the hospital and during a typical workday. As a result, the providers' experience with usability, and particularly the fit of the CDS tool within their workflow, may not mimic the true clinical setting. Unfortunately, it was not practical to perform usability testing in a true inpatient setting, given the nature of care in the hospital: providers are dealing with multiple patients, dealing with multiple issues, and working in multiple locations at any given time. As a result, hospital-based CDS systems are not triggered at an exact time or place in the workday and, therefore, real-time usability testing is only possible by shadowing providers around for many hours, which was not feasible in the context of this study. To assess usability in clinical practice, we plan to interview providers following implementation; nonetheless, interviews will have to occur after use rather than in real time because of these limitations in working in an inpatient setting. In addition, we plan to quantitatively measure the response rates by all providers for whom these alerts are triggered in a real-world clinical setting. Third, we interviewed a total of 12 providers. This number was based on reaching thematic

saturation, and previous studies have suggested 8 to 10 interviews to be sufficient for usability testing [15]. However, our sample was insufficient to determine differences in provider responses by specialty, and our sample may not be representative of providers at NYU Langone Health. Fourth, this study focused on providers who actually place orders in the inpatient system. Attending physicians, although not placing orders, have a significant influence on the care plan and may also benefit from CDS interventions. We hypothesize that attending providers or consultants may have preferences for CDS formats that are similar to supervising residents, although this hypothesis requires further research.

Conclusions

In one of the first evaluations of comparative usability of interruptive and noninterruptive alerts, we found that there is a trade-off between optimizing visibility and limiting distractions from a current task for interruptive and noninterruptive versions of a CDS. Maximizing the fit of CDS into the workflow is a key element for usability. A potential approach to increase fit into workflow may be to target alert timing and format based on the individual provider role. Whether such an approach leads to an increased uptake in clinical practice needs further evaluation.

Acknowledgments

The authors thank Meg Ferrauiola for her assistance in development of the CDS tool. This study was supported by the Agency for Healthcare Quality and Research grant K08HS23683.

Conflicts of Interest

None declared.

References

1. Riggio JM, Sorokin R, Moxey ED, Mather P, Gould S, Kane GC. Effectiveness of a clinical-decision-support system in improving compliance with cardiac-care quality measures and supporting resident training. Acad Med 2009 Dec;84(12):1719-1726. [doi: 10.1097/ACM.0b013e3181bf51d6] [Medline: 19940579]
2. Qian Q, Manning DM, Ou N, Klarich MJ, Leutink DJ, Loth AR, et al. ACEi/ARB for systolic heart failure: closing the quality gap with a sustainable intervention at an academic medical center. J Hosp Med 2011 Mar;6(3):156-160. [doi: 10.1002/jhm.803] [Medline: 20652962]
3. Backman R, Bayliss S, Moore D, Litchfield I. Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence. Syst Rev 2017 Dec 13;6(1):255 [FREE Full text] [doi: 10.1186/s13643-017-0627-z] [Medline: 29237488]
4. Nanji K, Seger D, Slight S, Amato MG, Beeler PE, Her QL, et al. Medication-related clinical decision support alert overrides in inpatients. J Am Med Inform Assoc 2018 May 1;25(5):476-481. [doi: 10.1093/jamia/ocx115] [Medline: 29092059]
5. Payne T, Hines L, Chan R, Hartman S, Kapusnik-Uner J, Russ AL, et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. J Am Med Inform Assoc 2015 Nov;22(6):1243-1250. [doi: 10.1093/jamia/ocv011] [Medline: 25829460]
6. Horsky J, Phansalkar S, Desai A, Bell D, Middleton B. Design of decision support interventions for medication prescribing. Int J Med Inform 2013 Jun;82(6):492-503. [doi: 10.1016/j.ijmedinf.2013.02.003] [Medline: 23490305]
7. Phansalkar S, van der Sijs H, Tucker A, Desai AA, Bell DS, Teich JM, et al. Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. J Am Med Inform Assoc 2013 May 1;20(3):489-493 [FREE Full text] [doi: 10.1136/amiajnl-2012-001089] [Medline: 23011124]
8. Kucher N, Puck M, Blaser J, Bucklar G, Eschmann E, Lüscher TF. Physician compliance with advanced electronic alerts for preventing venous thromboembolism among hospitalized medical patients. J Thromb Haemost 2009 Aug;7(8):1291-1296 [FREE Full text] [doi: 10.1111/j.1538-7836.2009.03509.x] [Medline: 19522743]
9. Jenssen B, Shelov E, Bonafide C, Bernstein S, Fiks A, Bryant-Stephens T. Clinical decision support tool for parental tobacco treatment in hospitalized children. Appl Clin Inform 2016;7:399-411. [doi: 10.4338/ACI-2015-12-RA-0169] [Medline: 27437049]
10. Pevnick J, Li X, Grein J, Bell D, Silka P. A retrospective analysis of interruptive versus non-interruptive clinical decision support for identification of patients needing contact isolation. Appl Clin Inform 2013;4(4):569-582 [FREE Full text] [doi: 10.4338/ACI-2013-04-RA-0021] [Medline: 24454583]
11. Strom BL, Schinnar R, Bilker W, Hennessy S, Leonard CE, Pifer E. Randomized clinical trial of a customized electronic alert requiring an affirmative response compared to a control group receiving a commercial passive CPOE alert: NSAID--warfarin co-prescribing as a test case. J Am Med Inform Assoc 2010;17(4):411-415 [FREE Full text] [doi: 10.1136/jamia.2009.000695] [Medline: 20595308]
12. Press A, McCullagh L, Khan S, Schachter A, Pardo S, McGinn T. Usability testing of a complex clinical decision support tool in the emergency department: lessons learned. JMIR Hum Factors 2015 Sep 10;2(2):e14 [FREE Full text] [doi: 10.2196/humanfactors.4537] [Medline: 27025540]

13. Richardson S, Mishuris R, O'Connell A, Feldstein D, Hess R, Smith P, et al. "Think aloud" and "Near live" usability testing of two complex clinical decision support tools. Int J Med Inform 2017 Dec;106:1-8 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.06.003] [Medline: 28870378]
14. Clarke MA, Belden JL, Kim MS. Determining differences in user performance between expert and novice primary care doctors when using an electronic health record (EHR). J Eval Clin Pract 2014 Dec;20(6):1153-1161. [doi: 10.1111/jep.12277] [Medline: 25470668]
15. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
16. Kushniruk AW, Borycki EM, Kuwata S, Kannry J. Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments. Stud Health Technol Inform 2011;169:915-919. [doi: 10.3233/978-1-60750-806-9-915] [Medline: 21893879]
17. 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]
18. Blecker S, Katz SD, Horwitz LI, Kuperman G, Park H, Gold A, et al. Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data. JAMA Cardiol 2016 Dec 01;1(9):1014-1020 [FREE Full text] [doi: 10.1001/jamacardio.2016.3236] [Medline: 27706470]
19. Blecker S, Meisel T, Dickson VV, Shelley D, Horwitz LI. "We're Almost Guests in Their Clinical Care": Inpatient Provider Attitudes Toward Chronic Disease Management. J Hosp Med 2017 Dec;12(3):162-167 [FREE Full text] [doi: 10.12788/jhm.2699] [Medline: 28272592]
20. Yancy C, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Oct 15;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
21. Blecker S, Agarwal SK, Chang PP, Rosamond WD, Casey DE, Kucharska-Newton A, et al. Quality of care for heart failure patients hospitalized for any cause. J Am Coll Cardiol 2014 Jan 21;63(2):123-130 [FREE Full text] [doi: 10.1016/j.jacc.2013.08.1628] [Medline: 24076281]
22. Rose AF, Schnipper JL, Park ER, Poon EG, Li Q, Middleton B. Using qualitative studies to improve the usability of an EMR. J Biomed Inform 2005 Feb;38(1):51-60 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.006] [Medline: 15694885]
23. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
24. Hancock B, Windridge K, Ockleford E. NIHR Research Design Service for Yorkshire & the Humber. 2007. An introduction to qualitative research URL: https://www.rds-yh.nihr.ac.uk/wp-content/uploads/2013/05/5_Introduction-to-qualitative-research-2009.pdf [accessed 2019-03-20] [WebCite Cache ID 770ROdxpS]
25. Osheroff J. Improving Medication Use and Outcomes With Clinical Decision Support: Step-By-Step Guide. USA: Amer Soc of Health System; 2009.
26. Linder J, Rigotti N, Schneider L, Kelley JH, Brawarsky P, Schnipper JL, et al. Clinician characteristics and use of novel electronic health record functionality in primary care. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i87-i90 [FREE Full text] [doi: 10.1136/amiajnl-2011-000330] [Medline: 21900702]
27. McCullagh L, Sofianou A, Kannry J, Mann D, McGinn T. User centered clinical decision support tools: adoption across clinician training level. Appl Clin Inform 2014;5(4):1015-1025 [FREE Full text] [doi: 10.4338/ACI-2014-05-RA-0048] [Medline: 25589914]
28. Sittig D, Krall M, Dykstra R, Russell A, Chin H. A survey of factors affecting clinician acceptance of clinical decision support. BMC Med Inform Decis Mak 2006 Feb 1;6:6 [FREE Full text] [doi: 10.1186/1472-6947-6-6] [Medline: 16451720]
29. Austrian JS, Jamin CT, Doty GR, Blecker S. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. J Am Med Inform Assoc 2018 May 1;25(5):523-529. [doi: 10.1093/jamia/ocx072] [Medline: 29025165]
30. Levin A, Stevens PE, Bilous RW, Coresh J, de Francisco AL, de Jong PE, et al. Kidney disease : Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013;3(1):1-150 10.1038/kisup.2012.73.

Abbreviations

ACE: angiotensin-converting enzyme
CDS: clinical decision support
EHR: electronic health record
EF: ejection fraction
eGFR: estimated glomerular filtration rate

NP: nurse practitioner

Edited by A Kushniruk; submitted 10.10.18; peer-reviewed by M Patel, R Alturki, E Puigdomènech, N Robles, D Willett; comments to author 03.02.19; revised version received 17.02.19; accepted 27.02.19; published 17.04.19.

Please cite as:
Blecker S, Pandya R, Stork S, Mann D, Kuperman G, Shelley D, Austrian JS
Interruptive Versus Noninterruptive Clinical Decision Support: Usability Study
JMIR Hum Factors 2019;6(2):e12469
URL: http://humanfactors.jmir.org/2019/2/e12469/
doi:10.2196/12469
PMID:30994460

Original Paper

Anesthesiology Control Tower—Feasibility Assessment to Support Translation (ACTFAST): Mixed-Methods Study of a Novel Telemedicine-Based Support System for the Operating Room

Teresa Murray-Torres[1], MD; Aparna Casarella[2], BS; Mara Bollini[1], RN, BSN, MHA; Frances Wallace[3], MD; Michael S Avidan[1], MB BCh; Mary C Politi[4], PhD

[1]Department of Anesthesiology, Washington University School of Medicine in St. Louis, Washington University in St. Louis, St. Louis, MO, United States

[2]Brown School of Social Work, Washington University in St. Louis, St. Louis, MO, United States

[3]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, United States

[4]Department of Surgery, Division of Public Health Sciences, Washington University School of Medicine in St. Louis, St. Louis, MO, United States

Corresponding Author:
Teresa Murray-Torres, MD
Department of Anesthesiology
Washington University School of Medicine in St. Louis
Washington University in St. Louis
660 S Euclid Avenue
Box 8054
St. Louis, MO,
United States
Phone: 1 314 226 8880
Fax: 1 314 454 2296
Email: murrayt@wustl.edu

Abstract

Background: Despite efforts to improve patient outcomes, major morbidity and mortality remain common after surgery. Health information technologies that provide decision support for clinicians might improve perioperative and postoperative patient care. Evaluating the usability of these technologies and barriers to their implementation can facilitate their acceptance within health systems.

Objective: This manuscript describes usability testing and refinement of an innovative telemedicine-based clinical support system, the Anesthesiology Control Tower (ACT). It also reports stakeholders' perceptions of the barriers and facilitators to implementation of the intervention.

Methods: Three phases of testing were conducted in an iterative manner. Phase 1 testing employed a think-aloud protocol analysis to identify surface-level usability problems with individual software components of the ACT and its structure. Phase 2 testing involved an extended qualitative and quantitative real-world usability analysis. Phase 3 sought to identify major barriers and facilitators to implementation of the ACT through semistructured interviews with key stakeholders.

Results: Phase 1 and phase 2 usability testing sessions identified numerous usability problems with the software components of the ACT. The ACT platform was revised in seven iterations in response to these usability concerns. Initial satisfaction with the ACT, as measured by standardized instruments, was below commonly accepted cutoffs for these measures. Satisfaction improved to acceptable levels over the course of revision and testing. A number of barriers to implementation were also identified and addressed during the refinement of the ACT intervention.

Conclusions: The ACT model can improve the standard of perioperative anesthesia care. Through our thorough and iterative usability testing process and stakeholder assessment of barriers and facilitators, we enhanced the acceptability of this novel technology and improved our ability to implement this innovation into routine practice.

International Registered Report Identifier (IRRID): RR2-10.1186/s40814-018-0233-4

(JMIR Hum Factors 2019;6(2):e12155) doi:10.2196/12155

KEYWORDS

clinical decision support systems; usability; anesthesiology; telemedicine

Introduction

The last several decades have brought major advancements in the safety of anesthetic techniques and therapeutics. However, patients undergoing surgery continue to experience persistent and significant risks of major morbidity and mortality following their operations [1]. Some of these risks are unavoidable and either inherent to the nature of the surgical procedure or attributable to patient characteristics not immediately modifiable [2-4]. However, many factors that impact a patient's immediate and long-term health can be influenced by the anesthesia care team [5-8]. Clinical decision support systems can optimize management of these factors, leading to improvement in intraoperative parameters such as hemodynamics [9], ventilator and fluid management [10], and blood glucose control [11,12]. Such systems are particularly useful for members of the anesthesia care team [13,14], who, like all medical practitioners, have known limitations in their cognitive abilities [15-17] and yet are often inundated with an overwhelming amount of information. Practitioners may see alarms as frequently as every 3 minutes and even more frequently during induction of anesthesia and emergence from anesthesia. Although the majority of alarms might appear clinically irrelevant, a small, critical percentage require immediate intervention [18]. Given the known limits of human cognitive abilities, there is a pressing need for decision support systems that improve clinicians' abilities to rapidly assess situations and act appropriately in a timely manner [13,14].

Decision support systems provide an opportunity to impact provider behaviors and patient outcomes in a broad range of clinical settings [19-21]. However, these interventions may fail to meaningfully influence patient care if they are not acceptable to the relevant end users [22]. Successful systems are those that achieve high levels of usability [23,24] by meeting thresholds for efficiency, effectiveness, and satisfaction [25-27] in the actual environment in which they will be used [22,28,29]. However, even a usable, well-designed intervention may fail if barriers to its integration into existing workflow patterns have not been considered [29,30]. The assessment of such barriers is particularly important in the setting of clinical trials [31,32], in which the delivery of an intervention is dependent on changes in behavior across many groups of individuals working in complex systems.

At our institution, we developed a novel telemedicine-based decision support intervention for the operating room called the Anesthesiology Control Tower (ACT), described in Multimedia Appendix 1 and elsewhere [33]. In the ACT, clinicians use several electronic health records (EHRs) to monitor surgical patients in real time and respond to clinical alerts generated by a customized version of a decision support software device called AlertWatch (AlertWatch LLC). This software system, modified in response to the testing described in this manuscript to create an AlertWatch Tower Mode, is a monitoring and alerting program that integrates information from patient monitors and EHRs. After analyzing the data, the program determines and displays the current patient state and generates alerts based on the incoming variables (Multimedia Appendix 1). A key component of the ACT is the presence of trained clinicians who are able to process these alerts. Just as an air traffic control tower monitors individual aircraft and delivers additional information and alerts to the pilot and copilot, the ACT functions as a clinical support system for teams of anesthesia clinicians, engaging with them to assist in providing safe, effective, and efficient care for their patients [34]. The ACT is currently being evaluated in the form of a proof-of-principle pragmatic trial [NCT02830126] [35].

The successful execution of complex health interventions such as the ACT demands an understanding of the usability of the intervention and any barriers and facilitators to its acceptance. Therefore, we designed this study to evaluate the ACT from the perspective of two groups of key stakeholders: those who deliver and those who receive the ACT support. Specifically, we aimed to determine whether the ACT adequately addressed goals for functionality and usability for end users. We also sought to identify barriers and facilitators to implementation of the ACT into routine clinical practice. We used these findings to modify the ACT based on user feedback.

Methods

Study Design

Three phases of testing were designed to determine the extent to which the different aspects of the ACT prototype met the needs of end users (see Table 1). A full description of the study protocol was previously published by our group [33]. Two phases of pragmatic mixed-method usability analyses [36-40] were conducted with ACT clinicians. These phases evaluated the entirety of the ACT structure and its software components. In the third phase of testing, semistructured interviews were conducted with operating room (OR) clinicians to identify barriers and facilitators to implementation of the ACT and obtain basic usability data. The results from testing phases were intended to guide iterative changes to the ACT structure and software, in particular the AlertWatch Tower Mode platform. Decisions to modify any component of the pilot ACT during the testing period were determined by the investigative team through review of participant feedback.

Table 1. Description of testing stages.

Stage of testing	Description	Eligible participants[a]	Outcome measures
Phase 1	Structured think-aloud usability sessions with ACT[b] clinicians	• Attending anesthesiologists • Resident anesthesiologists	• Task performance data • Standardized questionnaires • Utterance data
Phase 2	Near-live usability testing with ACT clinicians	• Attending anesthesiologists • Resident anesthesiologists	• Task performance data • Standardized questionnaires
Phase 3	Semistructured interviews with operating room clinicians	• Attending anesthesiologists • Resident anesthesiologists • Certified registered nurse anesthetists	• Barriers and facilitators to implementation

[a]Only physician anesthesiologists were eligible for participation in phase 1 and phase 2 based on the preliminary staffing model for the ACT.

[b]ACT: Anesthesiology Control Tower.

Participants

Participants were recruited from a single academic medical center using standardized emails distributed to the departmental listserv. All participants completed informed consent prior to study activities, consistent with the protocol approved by the Washington University in St. Louis Institutional Review Board (IRB #201611035). The target sample size for each round of testing was 8 to 10 participants based on guidelines for cognitive usability testing [41]. Based on an initial, physician-only staffing model for the ACT, only attending and resident physician anesthesiologists were eligible for participation in phase 1 and phase 2. All OR clinicians (physician anesthesiologists and certified registered nurse anesthetists [CRNAs]) were eligible to participate in phase 3.

Study Procedures

Usability Testing

Phase 1 was an exploratory think-aloud usability analysis with the two groups of clinicians who were eligible to staff the ACT (ACT clinicians). It aimed to identify major surface-level usability problems with the different components of the ACT [42], including orientation and help documents prepared by the research team, individual software components, and physical equipment and layout. A research team member was present in the room to moderate each session. Participants had 20 minutes to load the AlertWatch Tower Mode software in addition to the hospital's standard perioperative software programs, including the general EHR and the anesthesia information management system. They were instructed to address as many AlertWatch Tower Mode alerts as they could while voicing their thoughts and actions aloud [43]. Participants were prompted if 20 seconds elapsed without verbalization. If participants experienced a critical usability problem that prevented the session from continuing, the moderator provided the minimum amount of prompting that allowed the session to proceed. Sessions ended with structured debriefing sessions.

All think-aloud sessions were audio recorded and transcribed manually by a professional transcription service. Debriefing sessions were also transcribed when a recording was available. Transcripts were verified by the research team. At the end of each session, participants completed the quantitative usability and workload measures System Usability Scale (SUS) [44], Computer System Usability Questionnaire (CSUQ) [45], and NASA Task Load Index (NASA-TLX) [46].

Phase 2 consisted of usability testing of the ACT within its intended, real-world setting. No study personnel were present for this testing. Participants used the same suite of software programs as in Phase 1 to monitor surgical patients in real time and address the AlertWatch Tower Mode alerts. They did not interact with clinicians in the ORs. A secure server captured and stored a log of all alerts and participant responses. Testing sessions were of one business day duration. Based on the anticipated staffing model for the intervention, attending anesthesiologists participated one day at a time on days in which they were not assigned to the surgical ORs. Resident anesthesiologists participated for 10 consecutive business days as part of a formal 2-week rotation during their final year of clinical training. All participants completed the SUS [44], CSUQ [45], and NASA-TLX [46]. Attending anesthesiologists completed the questionnaires every day that they were in the ACT. To minimize the degree to which resident anesthesiologists were biased by their previous questionnaire responses, they only completed the surveys 3 times over the course of their rotations rather than on a daily basis.

Clinician Interviews

Phase 3 involved semistructured interviews with clinicians who were potential recipients of ACT feedback. After an initial orientation to the ACT, participants were prompted to provide their initial impressions of the intervention. Subsequently, a research team member presented six examples of scenarios for which participants were instructed to imagine themselves as the actual recipient of ACT feedback in each scenario. Five of the scenarios involved clinical alerts; the final scenario included a billing alert. After participants verbalized their initial reaction, the team member used a short series of open-ended questions to obtain input about the usefulness of each individual alert as well as the preferred delivery mode (eg, text, page, phone call). A debriefing session used open-ended questions to ascertain participants' final impressions of the ACT and their feedback on specific components of the ACT intervention. All interview sessions were audio recorded and professionally transcribed, with transcriptions verified against the original audio recordings.

Data Analysis

Quantitative Analysis

Participant characteristics from all three phases were analyzed using descriptive statistics. In phase 1, the frequency with which participants experienced a critical usability issue that required an intervention to continue the session was determined. Performance measures for the two ACT clinician usability testing phases were summarized with mean and standard deviation. These measures included time to task completion (phase 1) and quantity and rate of task completion (phases 1 and 2) [36,47]. For phase 2 testing, performance measures were analyzed across iterations of the software platform. Subjective measures of usability and workload in phases 1 and 2 were summarized with mean scores for the SUS and standard deviation for the CSUQ and NASA-TLX. Results from these surveys were compared between attending and resident physicians and between initial and repeat testing sessions. As a measure of participant satisfaction with the built-in software alerts, the percentage of alerts in phase 1 and phase 2 that were classified as significant or potentially significant was determined. Statistics were calculated using SPSS Statistics for Macintosh (IBM Corp).

Qualitative Analysis

Research team members used NVivo 12 software (QSR International Pty Ltd) to perform a qualitative content analyses of the ACT clinician think-aloud sessions and the OR clinician semistructured interviews. They analyzed transcripts in order to identify themes regarding the usability of the ACT (phase 1 ACT clinician sessions) and barriers and facilitators to its implementation (phase 3 OR clinician interviews). First, one researcher (TM-T) generated separate codebooks for each set of qualitative data (see Multimedia Appendix 2), and additional team members (AC, MB, and MP) helped refine them. Two researchers (TM-T and AC) double-coded transcripts until a kappa of at least .75 and percentage agreement of at least 97% had been obtained. Four transcripts from phase 1 and three transcripts from phase 3 were double coded. Subsequently these researchers coded the remaining transcripts independently. If consensus could not be reached during coding, a third team member (MP) reviewed the categorization.

Qualitative content analysis for phase 1 usability testing began with segmentation of the verbal data and ended with interpretation [48]. The final coding scheme (Multimedia Appendix 2) contained seven previously employed think-aloud content domains [49-54]. Analyses of the semistructured debriefing sessions focused on 2 of the 7 content domains (user experience, redesign proposal). In the final round of analyses, usability problems and redesign proposals were extracted from the transcripts [52,54]. Individual problems were placed into 1 of 4 categories (navigation, content, functionality, and layout; Multimedia Appendix 2). Redesign proposals were similarly grouped and compared with the themes from the usability problem set.

The analysis for phase 3 consisted of a qualitative usability analysis and thematic analysis. During this phase, one team member (TM-T) reviewed transcripts to explore participants' evaluation of the usefulness of the ACT in each individual clinical scenario in addition to the communication preferences of the participants. We based the thematic analysis on the theoretical domains framework that has demonstrated utility in examining constructs related to behavior change in a variety of health care settings [29,55-58]. After coding was complete, the researcher (TM-T) reviewed all utterances and created belief statements that captured meaningful themes within each domain [59]. These belief statements were summarized across participants and reviewed by two other team members (AC and MP).

Results

Participant Characteristics

Table 2 shows the characteristics of the participants at each stage of the testing process. A total of 32 clinicians participated over the course of all phases of testing. Eight attending and seven resident physician anesthesiologists participated in phase 1 testing; six attending and eight resident physician anesthesiologists participated in phase 2. Four attending physicians and six CRNAs participated in the semistructured OR clinician interviews. Resident physicians were also recruited for these interviews; however, due to challenges related to scheduling, no resident physicians were able to participate in the sessions.

Quantitative Data

Participants evaluated an average of 7.25 patients and addressed an average of 11.5 alerts per session during the phase 1 think-aloud usability sessions. During the phase 2 real-world testing, participants evaluated an average of 54.9 unique patients per day and addressed an average of 176 alerts across all platforms. Of the alerts addressed each day in phase 2, on average 40.3% (50/124) of them were repeat alerts—that is, one specific alert that triggered repeatedly for a single patient during a single operation.

The mean overall score for the SUS across all phase 1 and phase 2 sessions was 66.3, below the threshold of 70 that indicates a sufficient level of satisfaction [60]. The score tended to be higher during testing sessions that were repeat sessions versus the initial session (70 [SD 15] vs 62.6 [SD 16]) and for resident physicians versus attending physicians (70.3 [SD 14.9] vs 62.9 [SD 16.3]). Workload as measured by the NASA-TLX followed a similar pattern; lower workload was measured during repeat testing sessions (46.5 [SD 15.3] vs 53.2 [SD 18.3]) and among resident physicians versus attending physicians (50.1 [SD 12.9] vs 58.5 [SD 19.7]). No significant differences were found in CSUQ total or subscale scores between any testing conditions or participant roles. With regard to participant satisfaction with the specific software-generated alerts, participants determined that only 27.05% (680/2513) of the alerts generated by the first iteration of the platform were actually clinically significant or potentially significant. In subsequent iterations of the platform, this proportion of clinically useful alerts improved to more than half to three-quarters of all software-generated alerts (range 56.00% [933/1666] to 73.05% [1640/2245]).

Table 2. Characteristics of participant groups.

Characteristics	Attending anesthesiologists (n=16)	Resident anesthesiologists (n=10)	Certified registered nurse anesthetists (n=6)
Participants at each phase[a] (n)			
Phase 1 ACT[b] clinician think-aloud sessions	8	7	N/A[c]
Phase 2 ACT clinician real-world testing	6	8	N/A
Phase 3 OR[d] clinician interviews	4	0	6
Years at institute, average (range)	6.6 (0.75-21)	11.8 (3-22)	3.9 (3-4)
Sex, n (%)			
Male	10 (63)	5 (50)	3 (50)
Female	6 (38)	5 (50)	3 (50)
Baseline AlertWatch use, n (%)			
Almost always	1 (6)	2 (25)	1 (17)
Sometimes	6 (38)	3 (33)	2 (33)
Rarely or never	8 (50)	2 (25)	3 (50)

[a]There was an overlap of nine participants between phase 1 and phase 2 and two participants between phase 2 and phase 3. No participants overlapped between phase 1 and phase 3.

[b]ACT: Anesthesiology Control Tower.

[c]N/A: not applicable.

[d]OR: operating room.

Qualitative Data

Usability Problems

A total of 155 usability problems were identified in the phase 1 transcripts (Table 3), the majority of which were related to functionality (57/155, 36.8%) and content (53/155, 34.2%), followed by navigation (29/155, 18.7%) and layout (16/155, 10.3%). Three participants experienced a critical usability issue that required an intervention on the part of the research team member in order to continue in the session. Two of these participants were unable to locate the help documents that provided instructions on how to access the software programs, and the third participant loaded the wrong software platform. All three participants were able to continue in their sessions once the researcher pointed out the location of the help documents on the computer desktop.

The remaining usability problems were associated with delays in task performance or had minor effects on the testing session. Many users had difficulty both in understanding the meaning or relevance of the software-generated alerts (8/15, 53%) and determining the alert severity or priority (7/15, 46%). Some (7/15, 46%) noted being distracted by what they viewed as irrelevant or repeated alerts, and a few (3/15, 20%) reported being overwhelmed by the sheer number of alerts that they faced. Most clinicians reported limitations in their ability to address alerts and monitor patients due to poor interoperability and lack of integration of the different software programs (6/15, 40%) as well as slow response times for the software programs themselves (7/15, 46%). The fewest usability problems were associated with layout and were focused on participants' inability to move applications to the preferred of three monitors (3/15, 20%) or resize the program windows (2/15, 20%).

Redesign Proposals

Redesign proposals or suggestions for improvements were described by 12 of the 15 participants, with a total of 51 proposals reported from all participants. The proposals were often associated with the usability problems that participants encountered. Examples of these redesign proposals are provided in Multimedia Appendix 3. The majority of proposals were related to content (32/51, 63%) and included suggestions for improving alert relevance (13/51, 25%) and alert prioritization (9/51, 18%). Seven iterations of the AlertWatch Tower Mode platform (described in Multimedia Appendix 4) were tested during the trial and reflected the proposals generated by participants. The alterations included refinements in the visual display and presentation of information and changes in alert content and prioritization (shown in Multimedia Appendix 3).

Operating Room Clinician Perspective on Usability

In phase 3, the five clinical scenarios presented to participants were considered to be useful or potentially useful by all (10/10, 100%, for two of the scenarios) or almost all (8-9/10, 80%-90%, for the remaining scenarios) of the participants. Clinicians often had suggestions for how the usefulness of alerts could be improved, and many offered additional scenarios in which they would be satisfied with the usefulness of the ACT. Participants generally agreed that the preferred method of contact would depend on the clinical scenario; minor alerts could be sent through text, page, or even through the creation of a novel computer pop-up, whereas major alerts could be delivered by phone. The general consensus was that in order for a method of communication to be useful, it could not increase the provider's workload, distract from their current tasks, or add to their alarm fatigue.

Table 3. Usability problems identified in the Anesthesiology Control Tower clinician think-aloud and debriefing sessions.

Category[a] and theme	Number reporting	Example quotation
Navigation		
Trouble finding link or information	5	"Okay, so I have already forgotten what the heck I'm supposed to do to respond. I need to get that thing where I can click on 'responses' and I don't remember where it is." [Participant 2127, attending physician]
Unable to determine which link to use	2	"I don't know the difference between [two log-in options]. I don't know which one to do." [Participant 2108, attending physician]
Selected incorrect patient or operating room	2	"So here I was accidentally using the last patient we had, looking at that patient, before I realized that I was not on the correct patient." [Participant 2114, attending physician]
Any navigation problem	9	—[b]
Content		
Alert meaning or relevance unclear	8	"I'm unclear as to what infusions 4.0 means—whether that means 4 different types of infusions? ...I'm not sure what this means." [Participant 2105, resident physician]
Difficulty prioritizing alerts	7	"Which is worse, black or red? I'm guessing red...that wasn't spelled out to me, but I'm going to say yes." [Participant 2127, attending physician]
Information not available	6	"We are basically looking at a blank sheet with blood pressures randomly listed. I am unable to make any sort of reasonable clinical judgment at this point." [Participant 2114, attending physician]
Unable to identify correct patient or operating room	4	"What OR[c] is this again? I forgot what OR it is." [Participant 2112, resident physician]
Any dialogue problem	11	—
Functionality		
Poor software response times	7	"Waiting for [anesthesia information management system] to log in. Still waiting." [Participant 2127, attending physician]
Limited interoperability of software programs	7	"I'm a little frustrated because right now it seems kind of a hassle to access all these programs to make a simple decision." [Participant 2110, attending physician]
Inability to manipulate location of software programs on screen	3	"[The anesthesia information management system] won't let me move it to another screen. Looks like that is stuck on my middle screen, where [the EHR] I was able to move from monitor to monitor." [Participant 2106, attending physician]
Difficulty logging in to programs	5	"For some reason it does not allow me to log in or use [hospital] access." [Participant 2101, attending physician]
Any functionality problem	13	—
Layout		
Text not visible	4	"I'll have to spend a minute here trying to cover my cursor over to read the full case...chest wall reconstruction. It's sort of hard because it keeps going away." [Participant 2114, attending physician]
AlertWatch does not fit	2	"I first noticed AlertWatch is off the screen a little bit, trying to see if I can make it fit better—it doesn't really fit." [Participant 2103, attending physician]
Physical layout (monitors)	3	"How do I get the big board on the big screen? On the right? By convention it should be on the left." [Participant 2103, attending physician]
Any layout problem	9	—

[a]Adapted from Zhao et al [54].

[b]Not applicable.

[c]OR: operating room.

Barriers to Implementation

The interviews with OR clinicians in phase 3 generated 33 summary belief statements (Multimedia Appendix 5). Of these belief statements, 20 addressed potential barriers to the ACT implementation. Several participants questioned the necessity of the ACT and whether there were other ways to support good clinical care that would not require a "control tower." Many could imagine themselves to be frustrated or annoyed with a poorly executed intervention (emotion). They reported that the usefulness of even well-designed and accurate alerts could be drastically limited if the alert were poorly timed and distracted the clinician from meaningful patient care tasks (beliefs about consequences).

Many participants viewed the ACT intervention as being in actual or potential conflict with their professional autonomy as clinicians. Several participants feared the downstream impact of the ACT on provider satisfaction and even the department's ability to recruit and retain talented personnel (social professional role and identity, social influences, beliefs about consequences). Some also expressed apprehension regarding the very concept of remote monitoring (social professional role and identity) and imagined that their colleagues would feel similarly (social influences). A few participants questioned how the ACT would integrate into the legal structure for the provision of anesthesia and whether it would disrupt existing relationships between members of the anesthesia care team (beliefs about consequences).

Attending physicians and CRNAs voiced concern that the ACT support would be redundant, and some doubted the ability of the ACT to provide useful information of which the provider was not already aware. Several clinicians also imagined that the ACT clinicians may not be able to understand a patient's comorbidities and anesthetic needs as well the primary team did themselves (memory, attention, decision processes). In this setting, they stated that the ACT could worsen their workload if they had to take additional time to provide the missing information that would have allowed the ACT to better understand the patient's situation (beliefs about consequences). Additionally, participants worried that current limitations or flaws in monitoring and software systems could lead to the generation of false alarms or prevent clinicians from being able to act meaningfully on the ACT support (environmental context and resources). Participants identified flaws in the communication modalities currently available at the hospital which they envisioned leading to impairments in the ACT's ability to deliver timely and useful information (environmental context and resources).

Facilitators of Implementation

Despite potential barriers to implementation, all of the clinicians were able to identify several specific instances in which they could see benefit from the ACT intervention. In general, participants agreed that a timely alerting system that did not increase their workload or interrupt patient care could be useful. Attending physicians stated that the ACT could be useful for them during times when they were covering multiple busy rooms, either notifying them of acute major events or of smaller, but still relevant, alerts in stable cases (social professional role and identity). The ACT was also thought to be particularly helpful for newly employed or inexperienced clinicians (social professional role and identity, beliefs about consequences). In true crises, participants stated that the ACT could be most useful in helping the OR clinician to obtain additional hands-on assistance or by reviewing electronic records for critically relevant information that the OR team could be missing in the midst of a dynamic clinical situation (memory, attention, decision processes).

Almost all of the clinicians agreed that the clinical practices described in the example scenarios were consistent with good anesthesia practice (knowledge, nature of behavior). Most reported that the concept of the ACT intervention was simple to understand (knowledge). Clinicians identified patient safety as a focus of their identity as an anesthesia provider and stated that any interventions that enhanced this would be welcome (social professional role and identity, beliefs about consequences). In contrast to the clinicians who were apprehensive regarding the concept of remote monitoring, some participants clearly expressed willingness to incorporate the intervention into practice at the hospital. One provider compared the ACT to telemedicine in the intensive care unit, reporting this as a positive factor in having another clinician watching out for them and the patient (optimism).

Discussion

Principal Findings

In this paper, we have described a thorough and iterative evaluation of a novel telemedicine-based intervention for the OR from the perspective of key groups of end users. Our findings related to usability problems and barriers to implementation are consistent with prior studies investigating the incorporation of novel information technologies into clinical practice, and they allowed us to refine aspects of the intervention prior to the initiation of a pragmatic randomized controlled trial. Previous studies have demonstrated the necessity of comprehensive usability testing before the implementation of health information technologies into routine practice. As one group of authors noted, "it would be unthinkable that the airline industry would have its first trial of an airplane's flight capabilities with real passengers" [61]. The usability testing that we performed in this study allowed us to "pilot" the ACT prior to its implementation, enabling us to identify and mitigate limitations arising from the technical aspects of the intervention.

Participants identified a number of surface-level usability problems during phase 1 usability think-aloud sessions [62]. Usability problems were often related to visual displays and software interfaces, limited availability of information, and poor interoperability of software programs, consistent with prior work introducing novel technologies into clinical practice [63,64]. The phase 2 real-world usability testing provided complementary insight into usability and workflow concerns in a realistic environment [65,66]. Participants identified similar concerns to those discussed during the think-aloud sessions, such as difficulty prioritizing alerts and receiving repetitive alerts for a single patient. The iterative changes that were made to the ACT intervention (Multimedia Appendix 3) improved

the interface and its interactive features [23]. The decrease in the number of insignificant alerts reduced participants' cognitive load and alarm fatigue [23], allowing them to focus more on addressing clinically meaningful situations.

Although many usability problems were addressed over the course of testing, the research team was unable to improve the communication between the individual software programs used by the participants. In order to understand the context of an alert generated by the AlertWatch Tower Mode software, participants almost always required supplementary information from the hospital EHR and the anesthesia information management system. This process required clinicians to manually load individual patients into each of these different software programs. This lack of a uniform system was a known limitation of the EHRs at our institution at the time this study was performed. Recently, however, our institution recently transitioned from multiple EHRs to a single, combined platform, and we anticipate that the change will address this limitation.

Semistructured interviews with potential recipients of the ACT support discovered a range of beliefs related to facilitators and barriers to implementation of the ACT similar to findings from previous studies [30,56,58,59]. Our participants identified a number of potential adverse consequences to the introduction of a new technology and set of processes into routine care [67], including concerns regarding increased work for clinicians, unfavorable workflow issues, untoward changes in communication patterns and practices, negative emotions, and unexpected changes in the power structure. Some of these barriers indicate reactance on the part of the clinicians; that is, they experienced negative reactions as a result of threats to their autonomy and freedom of choice [68]. In response to some of the barriers involving professional autonomy and the social roles and identities of the study participants, the final ACT intervention was modified from a physician-only staffing model to one that also incorporated the clinical expertise of certified and student nurse anesthetists. This staffing model reflects the current structure at our institution in which attending physicians, nurse anesthetists, and resident physicians play important and complementary roles in providing care to surgical patients.

Despite the number of barriers revealed during testing, participants identified several facilitators to implementation such as a cultural commitment to patient safety. Participants expressed a willingness to engage with the ACT in order to improve its applicability and usefulness in helping clinicians adhere to high standards for patient care. Suggestions regarding the timing and appropriateness of specific alerts were consistent with the research team's initial design for the ACT and were incorporated into the final intervention.

Limitations

Results of the study should be interpreted within the context of several study limitations. Participants worked in one academic center that may not be representative of all health care settings. Study participants in the different phases were not representative of the final staffing model for the ACT. Specifically, the initial, physician-only staffing model for the ACT led to the exclusion of CRNAs from phase 1 and phase 2 usability testing. Additionally, resident physician anesthesiologists did not participate in phase 3 due to scheduling conflicts with their daily assignments. The lack of input from CRNAs in phases 1 and 2 may have prevented us from identifying all relevant usability problems related to the ACT intervention. However, due to similar backgrounds and experiences with the hospital system's EHRs between the physician anesthesiologists and CRNAs, we anticipate that testing with the two groups might have identified similar usability problem themes. We did obtain feedback from CRNAs during the semistructured interviews regarding the usability of the ACT from the OR clinician perspective. Across all testing phases, selection bias may have resulted in participants who felt most strongly about the ACT intervention being more or less inclined to participate. Finally, no patient-related outcomes were included in this study [19]; however, an ongoing randomized controlled proof-of-principle trial at our institution is evaluating metrics of care quality and tracking patient outcomes.

Conclusions

This mixed-methods study explored concerns about incorporating an innovative telemedicine-based clinical support system into routine clinical practice. Consistent with recommendations for assessing complex health interventions prior to their implementation, this study conducted usability testing and an analysis of barriers to and facilitators of implementation based on a theoretical framework [31,32]. By assessing not only usability but also acceptability and relevance [22,32] for two groups of end users, we maximized the potential of the ACT intervention to provide clinicians with the right support, in the right format, at the right time in the care continuum [69], thereby enhancing the ability of the ACT to meaningfully impact patient care.

Acknowledgments

We would like to acknowledge members of the Anesthesiology Control Tower—Feasibility Assessment to Support Translation (ACTFAST) study group who participated in the design of the ACT, assisted in aspects of data collection, and provided input on modifications during this study: Stephen Gregory, Bradley A Fritz, Arbi Ben Abdallah, Daniel L Helsten, Troy S Wildes, Anshuman Sharma, Alex Kronzer, Sherry McKinnon, and Thaddeus Budelier.

This study received grant support from the Washington University Institute of Clinical and Translational Sciences Funding Program (award #4265). It has also received support from the Agency for Healthcare and Research Quality in preparation for the randomized controlled trial (R21 HS024581-01A1). The funding agencies had no role in the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Anesthesiology Control Tower description.

[PDF File (Adobe PDF File), 513KB - humanfactors_v6i2e12155_app1.pdf]

Multimedia Appendix 2

Phases 1 and 3 study session codebooks.

[PDF File (Adobe PDF File), 359KB - humanfactors_v6i2e12155_app2.pdf]

Multimedia Appendix 3

Summary of changes to the AlertWatch Tower Mode platform.

[PDF File (Adobe PDF File), 225KB - humanfactors_v6i2e12155_app3.pdf]

Multimedia Appendix 4

Redesign proposals offered by clinicians in phase 1.

[PDF File (Adobe PDF File), 116KB - humanfactors_v6i2e12155_app4.pdf]

Multimedia Appendix 5

Summary of relevant belief statements and sample quotes from phase 3 clinicians.

[PDF File (Adobe PDF File), 97KB - humanfactors_v6i2e12155_app5.pdf]

References

1. Bainbridge D, Martin J, Arango M, Cheng D, Evidence-based Peri-operative Clinical Outcomes Research (EPiCOR) Group. Perioperative and anaesthetic-related mortality in developed and developing countries: a systematic review and meta-analysis. Lancet 2012 Sep 22;380(9847):1075-1081. [doi: 10.1016/S0140-6736(12)60990-8] [Medline: 22998717]
2. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 1999 Sep 07;100(10):1043-1049. [Medline: 10477528]
3. Turrentine FE, Wang H, Simpson VB, Jones RS. Surgical risk factors, morbidity, and mortality in elderly patients. J Am Coll Surg 2006 Dec;203(6):865-877. [doi: 10.1016/j.jamcollsurg.2006.08.026] [Medline: 17116555]
4. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013 Nov;217(5):833-842 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2013.07.385] [Medline: 24055383]
5. Kheterpal S, Tremper KK, Englesbe MJ, O'Reilly M, Shanks AM, Fetterman DM, et al. Predictors of postoperative acute renal failure after noncardiac surgery in patients with previously normal renal function. Anesthesiology 2007 Dec;107(6):892-902. [doi: 10.1097/01.anes.0000290588.29668.38] [Medline: 18043057]
6. Aronson S, Stafford-Smith M, Phillips-Bute B, Shaw A, Gaca J, Newman M, Cardiothoracic Anesthesiology Research Endeavors. Intraoperative systolic blood pressure variability predicts 30-day mortality in aortocoronary bypass surgery patients. Anesthesiology 2010 Aug;113(2):305-312. [doi: 10.1097/ALN.0b013e3181e07ee9] [Medline: 20571360]
7. Biccard BM, Rodseth RN. What evidence is there for intraoperative predictors of perioperative cardiac outcomes? A systematic review. Perioper Med (Lond) 2013 Jul 03;2(1):14 [FREE Full text] [doi: 10.1186/2047-0525-2-14] [Medline: 24472327]
8. Walsh M, Devereaux PJ, Garg AX, Kurz A, Turan A, Rodseth RN, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. Anesthesiology 2013 Sep;119(3):507-515. [doi: 10.1097/ALN.0b013e3182a10e26] [Medline: 23835589]
9. Nair BG, Horibe M, Newman S, Wu W, Peterson GN, Schwid HA. Anesthesia information management system-based near real-time decision support to manage intraoperative hypotension and hypertension. Anesth Analg 2014 Jan;118(1):206-214. [doi: 10.1213/ANE.0000000000000027] [Medline: 24247227]
10. Kheterpal S, Shanks A, Tremper KK. Impact of a novel multiparameter decision support system on intraoperative processes of care and postoperative outcomes. Anesthesiology 2018 Dec;128(2):272-282. [doi: 10.1097/ALN.0000000000002023] [Medline: 29337743]

11. Lipton JA, Barendse RJ, Schinkel AFL, Akkerhuis KM, Simoons ML, Sijbrands EJG. Impact of an alerting clinical decision support system for glucose control on protocol compliance and glycemic control in the intensive cardiac care unit. Diabetes Technol Ther 2011 Mar;13(3):343-349. [doi: 10.1089/dia.2010.0100] [Medline: 21291336]
12. Sathishkumar S, Lai M, Picton P, Kheterpal S, Morris M, Shanks A, et al. Behavioral modification of intraoperative hyperglycemia management with a novel real-time audiovisual monitor. Anesthesiology 2015 Jul;123(1):29-37. [doi: 10.1097/ALN.0000000000000699] [Medline: 26001031]
13. Kruger GH, Tremper KK. Advanced integrated real-time clinical displays. Anesthesiol Clin 2011 Sep;29(3):487-504. [doi: 10.1016/j.anclin.2011.05.004] [Medline: 21871406]
14. Rothman B, Sandberg WS, St Jacques P. Using information technology to improve quality in the OR. Anesthesiol Clin 2011 Mar;29(1):29-55. [doi: 10.1016/j.anclin.2010.11.006] [Medline: 21295751]
15. Dawson NV, Arkes HR. Systematic errors in medical decision making: judgment limitations. J Gen Intern Med 1987;2(3):183-187. [Medline: 3295150]
16. Stiegler MP, Ruskin KJ. Decision-making and safety in anesthesiology. Curr Opin Anaesthesiol 2012 Dec;25(6):724-729. [doi: 10.1097/ACO.0b013e328359307a] [Medline: 23128454]
17. Stiegler MP, Tung A. Cognitive processes in anesthesiology decision making. Anesthesiology 2014 Jan;120(1):204-217. [doi: 10.1097/ALN.0000000000000073] [Medline: 24212195]
18. de Man FR, Greuters S, Boer C, Veerman DP, Loer SA. Intra-operative monitoring—many alarms with minor impact. Anaesthesia 2013 Aug;68(8):804-810 [FREE Full text] [doi: 10.1111/anae.12289] [Medline: 23745968]
19. Sahota N, Lloyd R, Ramakrishna A, Mackay JA, Prorok JC, Weise-Kelly L, et al. Computerized clinical decision support systems for acute care management: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011;6:91 [FREE Full text] [doi: 10.1186/1748-5908-6-91] [Medline: 21824385]
20. Morris AH, Hirshberg E, Sward KA. Computer protocols: how to implement. Best Pract Res Clin Anaesthesiol 2009 Mar;23(1):51-67. [Medline: 19449616]
21. Richardson W, Berwick D, Bisgard J, Bristow L, Buck C, Cassel C. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington: National Academy Press; 2001.
22. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Dec 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
23. Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. J Am Med Inform Assoc 2018 May 01;25(5):585-592. [doi: 10.1093/jamia/ocx118] [Medline: 29126196]
24. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [Medline: 11922931]
25. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Software Quality Journal 2003;11(4):325-338. [doi: 10.1023/A:1025869312943]
26. Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: guidance on usability. Geneva: International Organization for Standardization; 1998.
27. Hornbæk K, Law E. Meta-analysis of correlations among usability measures. 2007 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: ACM; 2007; San Jose. [doi: 10.1145/1240624.1240722]
28. Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. BMJ 2002 Oct 26;325(7370):941 [FREE Full text] [Medline: 12399345]
29. Presseau J, Mutsaers B, Al-Jaishi A, Squires J, McIntyre C, Garg A, Major Outcomes with Personalized Dialysate TEMPerature (MyTEMP) Investigators. Barriers and facilitators to healthcare professional behaviour change in clinical trials using the Theoretical Domains Framework: a case study of a trial of individualized temperature-reduced haemodialysis. Trials 2017 Dec 22;18(1):227 [FREE Full text] [doi: 10.1186/s13063-017-1965-9] [Medline: 28532509]
30. Tavender EJ, Bosch M, Gruen RL, Green SE, Knott J, Francis JJ, et al. Understanding practice: the factors that influence management of mild traumatic brain injury in the emergency department--a qualitative study using the Theoretical Domains Framework. Implement Sci 2014 Jan 13;9:8 [FREE Full text] [doi: 10.1186/1748-5908-9-8] [Medline: 24418161]
31. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
32. O'Cathain A, Thomas KJ, Drabble SJ, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. BMJ Open 2013 Jun 20;3(6) [FREE Full text] [doi: 10.1136/bmjopen-2013-002889] [Medline: 23794542]
33. Murray-Torres TM, Wallace F, Bollini M, Avidan MS, Politi MC. Anesthesiology Control Tower: Feasibility Assessment to Support Translation (ACT-FAST)—a feasibility study protocol. Pilot Feasibility Stud 2018;4:38 [FREE Full text] [doi: 10.1186/s40814-018-0233-4] [Medline: 29416871]

34. Lobach D, Sanders GD, Bright TJ, Wong A, Dhurjati R, Bristow E, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. Evid Rep Technol Assess (Full Rep) 2012 Apr(203):1-784. [Medline: 23126650]
35. Gregory S, Murray-Torres TM, Fritz BA, Ben Abdallah A, Helsten DL, Wildes TS, ACTFAST Study Group. Study protocol for the Anesthesiology Control Tower—Feedback Alerts to Supplement Treatments (ACTFAST-3) trial: a pilot randomized controlled trial in intraoperative telemedicine. F1000Res 2018;7:623 [FREE Full text] [doi: 10.12688/f1000research.14897.2] [Medline: 30026931]
36. Hornbæk K. Current practice in measuring usability: challenges to usability studies and research. Int J Hum-Comput Stud 2006 Feb;64(2):79-102. [doi: 10.1016/j.ijhcs.2005.06.002]
37. Belden J, 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: https://www.himss.org/file/1268216/download?token=L4oSWWi1 [accessed 2019-03-16]
38. Daniels J, Fels S, Kushniruk A, Lim J, Ansermino JM. A framework for evaluating usability of clinical monitoring technology. J Clin Monit Comput 2007 Oct;21(5):323-330. [doi: 10.1007/s10877-007-9091-y] [Medline: 17701381]
39. Zahabi M, Kaber DB, Swangnetr M. Usability and safety in electronic medical records interface design: a review of recent literature and guideline formulation. Hum Factors 2015 Aug;57(5):805-834. [doi: 10.1177/0018720815576827] [Medline: 25850118]
40. Seffah A, Donyaee M, Kline R, Padda H. Usability measurement and metrics: a consolidated model. Software Qual J 2006 Jun;14(2):159-178. [doi: 10.1007/s11219-006-7600-8]
41. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
42. Rubin J, Chisnell D. Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests. Somerset: John Wiley & Sons; 2008.
43. Ericsson K, Simon H. Verbal reports as data. Psychol Rev 1980;87(3):215-251. [doi: 10.1037/0033-295X.87.3.215]
44. Brooke J. SUS: a quick and dirty usability scale. In: Jordan P, Thomas B, Weerdmeester B, McClelland I, editors. Usability Evaluation in Industry. Bristol: Taylor and Frances; 1996:4-7.
45. Lewis J. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
46. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 1988;52:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
47. Bevan N, Macleod M. Usability measurement in context. Behav Inform Technol 1994 Jan;13(1-2):132-145. [doi: 10.1080/01449299408914592]
48. Chi M. Quantifying qualitative analyses of verbal data: a practical guide. J Learning Sci 1997 Jul 1;6(3):271-315. [doi: 10.1207/s15327809jls0603_1]
49. Cooke L. Assessing concurrent think-aloud protocol as a usability test method: a technical communication approach. IEEE Trans Profess Commun 2010 Sep;53(3):202-215. [doi: 10.1109/TPC.2010.2052859]
50. Eveland JW, Dunwoody S. Examining information processing on the World Wide Web using think aloud protocols. Media Psychol 2000 May;2(3):219-244. [doi: 10.1207/S1532785XMEP0203_2]
51. Hertzum M, Borlund P, Kristoffersen K. What do thinking-aloud participants say? A comparison of moderated and unmoderated usability sessions. Int J Hum-Comput Interact 2015 Jun 26;31(9):557-570. [doi: 10.1080/10447318.2015.1065691]
52. McDonald S, Zhao T, Edwards H. Dual verbal elicitation: the complementary use of concurrent and retrospective reporting within a usability test. Int J Hum-Comput Interact 2013 Oct 03;29(10):647-660. [doi: 10.1080/10447318.2012.758529]
53. Zhao T, McDonald S. Keep talking: an analysis of participant utterances gathered using two concurrent think-aloud methods. 2010 Presented at: Proceedings of the 6th Nordic Conference on Human-Computer Interaction and extending Boundaries; 2010; Reykjavik. [doi: 10.1145/1868914.1868979]
54. Zhao T, McDonald S, Edwards H. The impact of two different think-aloud instructions in a usability test: a case of just following orders? Behav Inform Technol 2012 Aug 07;33(2):163-183. [doi: 10.1080/0144929X.2012.708786]
55. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
56. Curran JA, Brehaut J, Patey AM, Osmond M, Stiell I, Grimshaw JM. Understanding the Canadian adult CT head rule trial: use of the theoretical domains framework for process evaluation. Implement Sci 2013 Feb 21;8:25 [FREE Full text] [doi: 10.1186/1748-5908-8-25] [Medline: 23433082]
57. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
58. Patey AM, Islam R, Francis JJ, Bryson GL, Grimshaw JM, Canada PRIME Plus Team. Anesthesiologists' and surgeons' perceptions about routine pre-operative testing in low-risk patients: application of the Theoretical Domains Framework

(TDF) to identify factors that influence physicians' decisions to order pre-operative tests. Implement Sci 2012 Jun 09;7:52 [FREE Full text] [doi: 10.1186/1748-5908-7-52] [Medline: 22682612]
59. Francis JJ, Stockton C, Eccles MP, Johnston M, Cuthbertson BH, Grimshaw JM, et al. Evidence-based selection of theories for designing behaviour change interventions: using methods based on theoretical construct domains to understand clinicians' blood transfusion behaviour. Br J Health Psychol 2009 Nov;14(Pt 4):625-646. [doi: 10.1348/135910708X397025] [Medline: 19159506]
60. Bangor A, Kortum P, Miller J. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
61. Graham T, Kushniruk A, Bullard M, Holroyd B, Meurer D, Rowe B. How usability of a web-based clinical decision support system has the potential to contribute to adverse medical events. AMIA Annu Symp Proc 2008 Nov 06:257-261 [FREE Full text] [Medline: 18998968]
62. Jaspers MWM. 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]
63. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. JAMA 2018 Dec 27;319(12):1276-1278 [FREE Full text] [doi: 10.1001/jama.2018.1171] [Medline: 29584833]
64. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005 Aug;74(7-8):519-526. [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]
65. Li AC, Kannry JL, Kushniruk A, Chrimes D, McGinn TG, Edonyabo D, et al. Integrating usability testing and think-aloud protocol analysis with. Int J Med Inform 2012 Nov;81(11):761-772. [doi: 10.1016/j.ijmedinf.2012.02.009] [Medline: 22456088]
66. Kushniruk AW, Borycki EM, Kuwata S, Kannry J. Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments. Stud Health Technol Inform 2011;169:915-919. [Medline: 21893879]
67. Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc 2006;13(5):547-556 [FREE Full text] [doi: 10.1197/jamia.M2042] [Medline: 16799128]
68. Vashitz G, Meyer J, Parmet Y, Peleg R, Goldfarb D, Porath A, et al. Defining and measuring physicians' responses to clinical reminders. J Biomed Inform 2009 Apr;42(2):317-326 [FREE Full text] [doi: 10.1016/j.jbi.2008.10.001] [Medline: 19000935]
69. James BC. Making it easy to do it right. N Engl J Med 2001 Sep 27;345(13):991-993. [doi: 10.1056/NEJM200109273451311] [Medline: 11575294]

Abbreviations

ACT: Anesthesiology Control Tower
ACTFAST: Anesthesiology Control Tower—Feedback Alerts to Support Translation
CRNA: certified registered nurse anesthetist
CSUQ: Computer System Usability Questionnaire
EHR: electronic health record
NASA-TLX: NASA Task Load Index
OR: operating room
SUS: System Usability Scale

Edited by C Lovis; submitted 09.09.18; peer-reviewed by S Khairat, A Woodbury; comments to author 28.12.18; revised version received 21.02.19; accepted 13.03.19; published 23.04.19.

Please cite as:
Murray-Torres T, Casarella A, Bollini M, Wallace F, Avidan MS, Politi MC
Anesthesiology Control Tower—Feasibility Assessment to Support Translation (ACTFAST): Mixed-Methods Study of a Novel Telemedicine-Based Support System for the Operating Room
JMIR Hum Factors 2019;6(2):e12155
URL: http://humanfactors.jmir.org/2019/2/e12155/
doi:10.2196/12155
PMID:31012859

Original Paper

Evaluation of an Anesthesia Dashboard Functional Model Based on a Manufacturer-Independent Communication Standard: Comparative Feasibility Study

Marian Ohligs[1], MSc; Carina Pereira[1], MSc, Dr-Ing; Verena Voigt[1], MD, Dr med; Marcus Koeny[1], Dipl-Ing, Dr-Ing; Armin Janß[2], Dipl-Ing, Dr-Ing; Rolf Rossaint[1], Dr med; Michael Czaplik[1], PD, Dr rer nat, Dr med

[1]Department of Anesthesiology, Faculty of Medicine, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany

[2]Chair of Medical Engineering, Faculty of Mechanical Engineering, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany

Corresponding Author:
Marian Ohligs, MSc
Department of Anesthesiology, Faculty of Medicine
Rheinisch-Westfälische Technische Hochschule Aachen University
Pauwelsstrasse 30
Aachen, 52074
Germany
Phone: 49 241 80 ext 83136
Email: mohligs@ukaachen.de

Abstract

Background: Current anesthesia workspaces consist of several technical devices, such as patient monitors, anesthesia machines, among others. Commonly, they are produced by different manufacturers; thus, they differ in terms of their modus operandi, user interface, and representation of alarms. Merging the information from these devices using a single joint protocol and displaying it in a single graphical user interface could lead to a general improvement in perioperative management. For this purpose, the recently approved and published Institute of Electrical and Electronics Engineers 11073 service-oriented device connectivity standard was implemented.

Objective: This paper aims to develop and then evaluate an anesthesia workstation (ANWS) functional model in terms of usability, fulfillment of clinical requirements, and expected improvements in patient safety.

Methods: To compare the self-developed ANWS with the conventional system, a pilot observational study was conducted at the University Hospital Aachen, Germany. A total of 5 anesthesiologists were asked to perform different tasks using the ANWS and then the conventional setup. For evaluation purposes, response times were measured and an interaction-centered usability test with an eye-tracking system was carried out. Finally, the subjects were asked to fill in a questionnaire in order to measure user satisfaction.

Results: Response times were significantly higher when using the ANWS, but decreased considerably after one repetition. Furthermore, usability was rated as excellent (≥95) according to the System Usability Scale score, and the majority of clinical requirements were met.

Conclusions: In general, the results were highly encouraging, considering that the ANWS was only a functional model, as well as the lack of training of the participants. However, further studies are necessary to improve the universal user interface and the interplay of the various networked devices.

(JMIR Hum Factors 2019;6(2):e12553) doi:10.2196/12553

KEYWORDS

operating room; anesthesia; interconnection; networking; human-computer interaction; process optimization; intelligent alarms; decision-support systems; 11073 SDC; service-oriented device connectivity

Introduction

In recent decades, the development of new innovative medical devices has led to significant improvements in patient safety, quality of supply, and economic advantages. Conventional anesthesia workspaces, for instance, consist of many different technical devices, including (1) a patient monitor, which records and displays the patient's vital parameters, such as electrocardiogram, body temperature, blood pressure, oxygen saturation, pulse rate, and respiratory rate; (2) an anesthesia machine, which is used to support the administration of anesthesia and for mechanical ventilation; and (3) syringe pumps, which administer intravenous anesthesia. Nowadays, all of these devices are the basis for standard patient care in anesthesiology [1].

However, medical devices assembled for clinical applications are usually produced by different manufacturers. They vary in terms of their modus operandi, user interface, and representation of alarms [2,3]. In intensive care, which represents an example of a data-rich environment, studies have demonstrated that 80% of user errors are a result of cognitive overload [4]. Medical device interoperability allows the joining of all data sources and, thus, leads to a unified presentation and control. To date, there is a lack of a single, shared communication protocol as well as a common interface between devices from different manufacturers [5]. This is especially true among intelligent decision-support systems [6], monitoring systems, and supervision systems [7], which are becoming more and more relevant in modern clinical practice. The fusion of all of this important information (eg, patient data and system settings) in a single graphical user interface (GUI) would simplify the work of anesthesiologists and improve patient safety [8], for example, to better determine the depth of anesthesia and to better optimize perioperative management, including logistics [9,10]. In addition, current patient data management systems only collect information from a number of devices if a proprietary connector is available.

In order to solve the issues discussed, an interdisciplinary consortium composed of engineers, computer scientists, and physicians from approximately 50 German organizations (ie, research institutes, hospitals and clinics, and medical companies) initiated the Secure Dynamic Networking in the Operating Room and Clinic (OR.NET) research project, which was funded by the German Federal Ministry of Education and Research (grant number 16KT1238). Two of the main goals were to develop a single and sophisticated protocol—the service-oriented device connectivity (SDC) family of standards, formerly the Open Surgical Communication Protocol—for medical device communication, as well as a new anesthesia workstation (ANWS), focusing at the same time on human-computer interaction (HCI) and safety concepts [11]. The SDC standards were developed as a general peer-to-peer interconnection protocol for an accurate exchange of medical information (eg, vital parameters and alarms) within operating room networks. They were based on the Device Profile for Web Services standard [12,13], which ensures that each device can communicate in a service-oriented architecture with the help of standard Web services [14,15]. Standard Ethernet is used for system communication since it is cost-effective and supported by the vast majority of medical devices, allowing manufacturer-independent interconnections [11]. After several years of research, the developed OR.NET architecture was included in the Institute of Electrical and Electronics Engineers (IEEE) 11073-SDC family of standards (Health informatics—point-of-care medical device communication), which address interoperability of medical devices and clinical IT infrastructure [16].

In addition to technical interoperability issues, such as definition of data and communication protocols, safety and risk management is also part of the 11073-SDC communication protocol extensions 10207, 20702, and 20701. They were approved and published by the IEEE in January 2019. The so-called *safety classifications* regularize the responsibility among interconnected medical devices and software services.

The ANWS functional model was developed to demonstrate and examine the capabilities of medical device interconnections based on the SDC standards; the model is represented in Figure 1. Nonetheless, a common platform presents several advantages in terms of patient safety and teamwork in the operating theater. First, a GUI, which gathers the core information from different sources, offers the physician a better overview of the patient's overall state as well as of the whole clinical setup and setting. Secondly, interdisciplinary standard operating procedures (SOPs) and checklists (eg, the Surgical Safety Checklist [SSC]) integrated into the ANWS contribute to a better workflow between different medical departments, such as anesthesia and surgery [7,17]. Thirdly, a common platform might minimize the failure rate for documentation, while reducing the physician's workload as well as the number of nondigital documents [5].

The objective of this paper is to evaluate the ANWS functional model developed during the OR.NET research project in terms of usability, fulfillment of clinical requirements, and expected improvements in patient safety.

Figure 1. The anesthesia workstation (ANWS) functional model. The interface screen consists of six main elements: (a) patient context and alarms; (b) device panel: overview of all devices connected to the ANWS; (c) view mode selector; (d) area reserved for the selected view; (e) workflow process; and (f) eye-tracking system.

Methods

Study Design

Overview

In order to evaluate the usability of the OR.NET ANWS, and consequently the HCI, a study involving 5 professional volunteers was conducted at the University Hospital Aachen, Germany. Potential benefits of the functional model over a conventional system were examined. Therefore, every participant started with the OR.NET setting, followed by the conventional setup a few weeks later. During a total hip arthroplasty (ie, total hip replacement) surgery simulation, the usability of both systems was evaluated by measuring the usability criteria (ie, effectiveness, efficiency, learnability, and user satisfaction); criteria were measured by (1) using the response time and (2) applying the think-aloud method. Thinking aloud is often used in usability testing by asking the participants to say whatever comes into their mind while performing certain tasks. This method enables insights into the participant's cognitive processes, including perception, doing, and feeling [18,19]. For the evaluation of the ANWS, further factors based on eye tracking were examined, such as (1) detection of "vampire effects" (ie, when eye catchers draw away and consume the user's attention) [20]; (2) hidden affordance (ie, when the functionality and intended use of a particular control are not intuitive); and (3) areas of interest [18,19]. At the end of the study, the participants were asked to rate some important features of the ANWS, for example, automatic documentation, compilation of all alarms, and decision-support system.

Anesthesia Workstation

Overview

The ANWS was installed on a desktop computer with the following properties: Quad-Core i5 CPU, 8 GB RAM, 250 GB solid-state drive, and a Radeon R7 200 dedicated graphics card with 1 GB graphics double-data rate type 5 memory. It was started as a stand-alone executable based on the .NET Framework 4.6 and OpenStackClient C# library (OSCLib C#), version 0.97_09 (SurgiTAIX AG).

Simulator

An OR.NET-SDC device simulator (Ilara GmbH) was used during the study to simulate measurements and output data for the following medical devices: syringe pump, patient monitor, and anesthesia machine. Thus, predetermined clinical scenarios could be reproduced. The simulator transmitted scheduled numerical data (eg, heart rate, blood pressure, and oxygen saturation) as well as physiological waveforms (eg, electrocardiography and respiratory curves) inside the network. In addition, it was able to respond to external commands; these were used to set parameters in the simulated devices (eg, infusion rates of syringe pumps). It was started as a stand-alone executable based on Microsoft Office 2016, .NET Framework 4.5, and OSCLib C#, version 0.96_00. The ANWS and the simulator were connected using the SDC through an internal network on the test computer.

Eye Tracking

To monitor the users' behavior (ie, the users' visual attention on user interface elements), we used an eye-tracking camera system—Gazepoint GP3 HD with Gazepoint Analysis and Gazepoint Control recording software, version 3.5 (Gazept). This was installed below the external monitor, as displayed in Figure 1. Eye-tracking videos and eye-tracking measurements were recorded with approval from the participants. In addition, the desktop window was also acquired using the Gazepoint Analysis software.

Audio Acquisition

The subjects were asked to report or express every single thought while solving the tasks (ie, think-aloud method). During the study, an audio record was created for future analysis. Here, the microphone from the LIVE! Cam Chat HD webcam (Creative) was used.

Study Tasks

In the first phase, the participants had to carry out 42 tasks. As described in Table 1, a total of 33 tasks were identical for both study phases. To test additional features of the ANWS, nine further tasks were carried out. In general, they consisted of looking up a value displayed on the ANWS as well as setting new parameters, such as flow rate of syringe pumps and positive end-expiratory pressure (PEEP) of the anesthesia machine. Table 1 describes the tasks performed by the subjects during this usability study.

Questionnaire

At the end of the study, each participant filled in a survey that included questions about sociodemographic aspects, technical expertise, motivation, study conditions, and features tested. For each tested feature of the ANWS (see Table 2), the System Usability Scale (SUS) developed by John Brooke [21] was used to assess the user's opinion. To calculate the SUS, 10 specified questions were rated using a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Whereas the rates of the positively formulated questions were subtracted by 1, resulting in a score from 0 to 4, the rates of negatively formulated questions were subtracted from 5, also resulting in a score from 0 to 4. Finally, all 10 scores from the questions—five positively formulated and five negatively formulated—were summed and multiplied by 2.5. The results were graded from A to F and compared to acceptability ranges and adjective ratings, as illustrated in Figure 2 [21,22].

Table 1. Tasks carried out by the subjects in both study phases.

Task	Study phase	Number of repetitions
Select type of surgery and load respective workflow	ANWS[a]	1
Customize workstation window	ANWS	1
Update workflow process	ANWS	4
Read blood pressure	ANWS, C[b] (PM[c])	4
Read temperature	ANWS, C (PM)	3
Read oxygen saturation	ANWS, C (PM)	7
Read heart rate	ANWS, C (PM)	2
Read airway pressure	ANWS, C (AM[d])	4
Read fraction of inspired oxygen	ANWS, C (AM)	1
Read respiratory minute volume	ANWS, C (AM)	1
Read respiratory compliance	ANWS, C (AM)	1
Set positive end-expiratory pressure (ie, respirator parameter)	ANWS, C (AM)	2
Set infusion flow rate	ANWS, C (SP[e])	4
Fetch and complete Surgical Safety Checklist	ANWS, C	3
Consider a pulmonary embolism	ANWS, C	1
Switch workspace view	ANWS	2
Check for intelligent alarms	ANWS	1
Total tasks	ANWS, C	42 (ANWS); 33 (C)

[a]ANWS: anesthesia workstation.

[b]C: conventional setting.

[c]PM: patient monitor.

[d]AM: anesthesia machine.

[e]SP: syringe pump.

Table 2. Overview of all features integrated into the anesthesia workstation and respective descriptions.

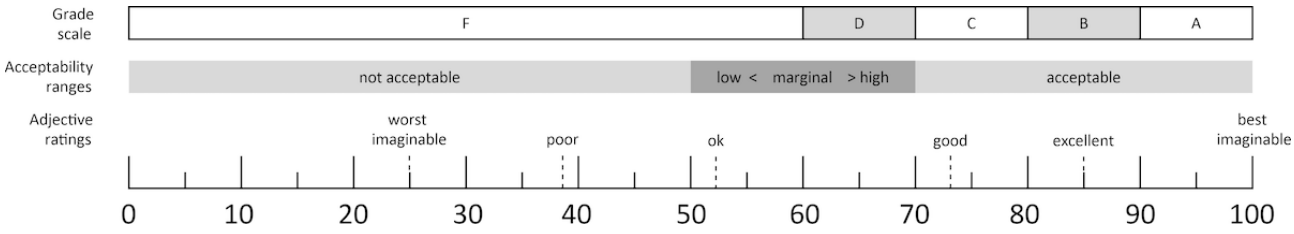
Feature[a]	Tested	Description
Automatic documentation	No	Possibility of saving all information provided by a service-oriented device connectivity-compatible device inside the OR.NET[b] network, including meta information, through the ANWS[c].
Compilation of all alarms	No	Alarms provided by any anesthesia-related device (eg, syringe pump, anesthesia machine, and patient monitor) are collected and displayed in the ANWS.
Control devices from other departments	No	Bidirectional control of devices (eg, anesthesiologists can control surgery devices and surgeons can control anesthesia equipment).
Display content from other departments	No	Capability to integrate several or even single measures or parameters in any connected device (eg, to display the current blood pressure, measured by the patient monitor, in the surgical microscope).
Cross-device interaction	Yes	Enables (eg, control of) diverse surgical devices using a single universal footswitch, button, or joystick.
Decision-support system	Yes	Context-adaptive hints and suggestions are displayed based on the currently ongoing surgical intervention, the actual workflow step, eventual patient-related problems, etc.
Segregated alarms	No	All alarms are classified as medical or device-associated (ie, technical) alarms.
Surgical Safety Checklist	Yes	Display of the Surgical Safety Checklist as an integrative part of the surgical workflow.
Unified surface	Yes	Fusion of information from several devices in a single graphical user interface.
Workflow management	Yes	Generation from a workflow based on previous data (ie, database). After each step, the process is updated automatically or manually, enabling significantly improved (ie, predictive) planning.

[a]All features were subjectively rated; five of those were additionally included in a user test scenario.

[b]OR.NET: Secure Dynamic Networking in the Operating Room and Clinic.

[c]ANWS: anesthesia workstation.

Figure 2. System Usability Scale (SUS) scores. Graphical overview of the adjective ratings, acceptability scores, and school grading scales, in relation to the average SUS score. Figure adapted from Bangor et al [22].



Finally, participants were asked to prioritize the tested features according to the relevance in their daily work. Furthermore, they were requested to compare the OR.NET ANWS with common or conventional systems in terms of personal preference and impact on patient safety.

Conventional System

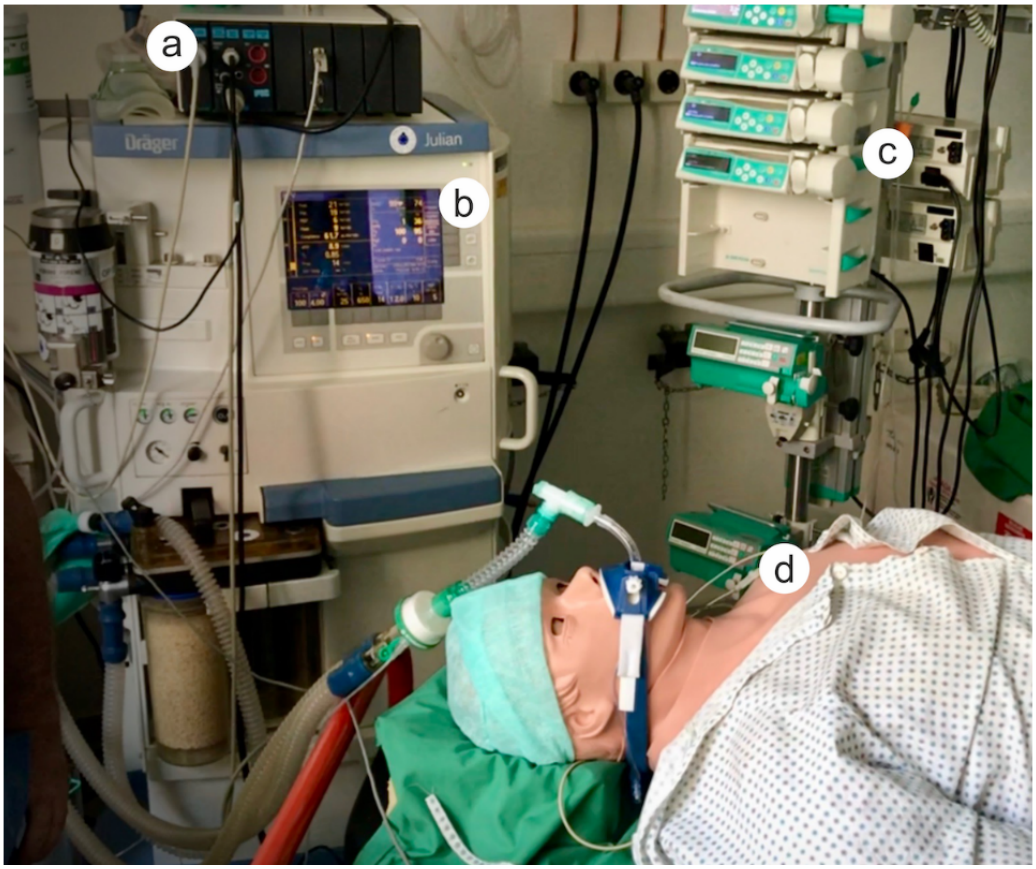
In the conventional phase, the subjects were asked to carry out 33 tasks (see Table 1). These consisted of looking up a value displayed on one of the devices (eg, blood pressure, temperature, and heart rate) or of setting one of these parameters (eg, flow rate of syringe pumps or PEEP). Note that these tasks were the same for the OR.NET and the conventional systems.

In this evaluation, participants used a Julian anesthesia machine (Draeger Medical), a Perfusor Space syringe pump (B Braun), a Datex-Ohmeda AS patient monitor (General Electric), and a filing folder including the printed SSCs for the three operation phases: (1) *sign in*: before induction of anesthesia; (2) *time out*: before incision of the skin; and (3) *sign out*: before the patient leaves the operating room. The SSCs contained the same items as those implemented in the ANWS workflow management. A high-end simulator room at the University Hospital Aachen (AIXTRA) was used to measure the time spans for the same participants carrying out the tasks using conventional machines [23]. For comparison purposes, the response times were measured. Figure 3 illustrates the study setup.

Study Participants

All 5 subjects—aged 30-42 years; 3 females (60%) and 2 males (40%)—participated in this study voluntarily. The group was composed of anesthesiologists—2 assistant physicians (40%) and 3 consultant anesthesiologists (60%)—from the Department of Anesthesiology of the University Hospital Aachen. The group was heterogeneous regarding practical experience: 3 physicians (60%) had more than 5 years of experience; 1 (20%) had 3-5 years of experience; and 1 (20%) had 1-2 years of experience. None of them had any knowledge of the study scenario or tasks. Only 1 (20%) of the participants had seen the GUI design of the ANWS before, without any connection to the study scenario or tasks. According to the guidelines of the ethics committee, no formal approval was necessary to conduct the study.

Figure 3. Illustration of the study setup: (a) patient monitor, (b) anesthesia machine, (c) syringe pump, and (d) patient simulator.



Data Analysis

The primary outcome parameters in this study were the response times and the capability of the subjects in successfully solving each task. As previously discussed, eye tracking was used in this study. The first aim was to identify elements that distracted attention from more important information, known as "vampire effect." The second aim was to find elements with a hidden affordance. Lastly, eye-tracking information was essential to find areas of interest as well as hitherto unknown impacts or suggestions for improvements.

Using the think-aloud method, it was possible to achieve a better understanding of the participant's line of thought, reasons for the occurrence of problems or misunderstandings, if applicable, and to identify potential further improvements to the system.

The questionnaire, in turn, was used to track background knowledge and motivation. Information on technical know-how and work experience was requested in order to determine the actual impact of each participant on the study, including a possible halo effect (ie, form of cognitive bias arising from an overgeneralization, such as a limited amount of evidence, the influence of preconceived beliefs, or a priori hypotheses; it is particularly prone to occur among participants who are enthusiastic about technology) [24]. With support from the questionnaire, SUS scores were calculated for the tested features: decision-support system, workflow management, SSC, customizable GUI, and cross-device interaction.

Statistical Analysis

Statistical analysis was performed using SPSS Statistics for Windows, version 22.0 (IBM Corp). Normal distribution was analyzed with Kolmogorov-Smirnov and Shapiro-Wilk tests. When detecting significance in the comparison of means (ie, between the conventional and OR.NET settings), the Mann-Whitney *U* test and the Wilcoxon signed-rank test were used. A *P* value <.05 was considered statistically significant.

Results

Questionnaire: Part A

Table 3 presents the results for the first part of the questionnaire. In general, all 5 participants were enthusiastic about new technologies and used them in their daily life. They classified their technical knowledge as normal or proficient. The great majority (4/5, 80%) agreed that they quickly become used to new technologies and that these make their lives easier. The subjects reported a daily use of mobile phones. The great majority used a tablet (3/5, 60%) and a PC or notebook (4/5, 80%) daily or at least several times a week. Interestingly, some users had already had some private and professional experience with other smart devices, such as smart watches and smart glasses.

Table 3. First part of the questionnaire.

Questions and responses	Score[a], median (min, max)
Technological use	
I often use technical innovation in everyday life.	4 (4, 4)
I'm skeptical of new technologies.	1 (1, 2)
I often use technical innovations to make my life easier.	4 (3, 4)
I quickly get used to using new technologies.	4 (3, 4)
I can easily use new technologies.	4 (3, 4)
I don't like surgical robots in medicine.	1 (1, 2)
In general, more technology should be used.	4 (2, 4)
New technologies endanger society.	1 (1, 2)
I use...	
A smartphone (private)	4 (4, 4)
A smartphone (work)	4 (4, 4)
A tablet (private)	3 (2, 4)
A tablet (work)	3 (2, 4)
A notebook or PC (private)	4 (1, 4)
A notebook or PC (work)	4 (1, 4)
Another device; private (eg, smart glasses or smart watch)	1 (1, 4)
Another device; work (eg, smart glasses or smart watch)	2 (1, 3)
Study participation	
Scientific studies do not support health care delivery and outcomes.	1 (1, 1)
I expect a compensation or reimbursement for expenses.	1 (1, 3)
Personally, I do not find questionnaires useful for gathering information on individual or collective perspectives.	1 (1, 3)
I participate in studies to learn.	3 (2, 4)
Test conditions	
I felt great pressure to perform.	1 (1, 3)
I felt excessively challenged.	1 (1, 1)
I was stressed.	2 (1, 2)
My motivation was high.	4 (3, 4)
I was focused.	4 (3, 4)

[a]Scores were as follows: 0 (*no statement*), 1 (*disagree*), 2 (*rather disagree*), 3 (*rather agree*), and 4 (*agree*).

All participants (5/5, 100%) agreed that such studies might improve health care delivery and outcomes. In fact, most of the volunteers wanted to participate in order to learn more (4/5, 80%) and did not expect any compensation or reimbursement for expenses (4/5, 80%). In addition, they considered the questionnaires to be helpful for gathering information on individual and collective perspectives.

Regarding the study itself, the participants did not feel pressured, stressed, or challenged. Instead, they were highly motivated and focused in trying to solve the tasks. In fact, 2 physicians (40%) had gained experience in telemedicine through the emergency medical service in Aachen, Germany [25,26]. In addition, 1 (20%) anesthesiologist was involved in configuring and implementing a decision-support system in an intensive care unit.

Response Time

One of the first tasks in this study consisted of finding the correct SOP. In the ANWS setting, the SOP was automatically generated by the decision-support engine with no delay and was embedded in the workflow management. However, in the conventional setting, the volunteers required 26.8 seconds to accomplish the task.

As described in Table 1, the subjects were asked in some cases to repeat the same task (ie, nonconsecutive repetitions). In the first round of questions, the time required to read out medical parameters was significantly lower when using the conventional setting (see Figure 4a). The response time was 1 second (interquartile range [IQR] 0-3) for the conventional setting and 2.5 seconds (IQR 0.75-4.25) for the ANWS setting (P=.04). A

significant decrease in response time was observed directly at the first repetition as depicted in the box plot of Figure 4a. In both phases, it was smaller than 1 second. Interestingly, no difference between groups (ie, phases) was found.

When the subjects were advised to set the PEEP in the anesthesia machine and in the workstation for the conventional and ANWS setting, respectively, the same phenomenon was observed (ie, they required less time when using the conventional approach: conventional, 2 seconds [IQR 2-3]; ANWS, 3 seconds [IQR 3-4]; *P*=.03). The box plot of Figure 4b compares both methods. In the first repetition, the response time for the conventional system did not change: 2 seconds (IQR 2-3). In turn, a decrease in response time for the ANWS setting was observed: 2 seconds (IQR 1-2). Here, the *P* value was .52. In this case, a learning effect with decreasing response time was observed for the ANWS: 3 seconds (IQR 3-4) versus 2 seconds (IQR 1-2; *P*=.35). This was not observed for the conventional setting: 2 seconds (IQR 2-3) versus 2 seconds (IQR 2-3; *P*=.66).

As represented in Figure 4c, the same trend was not visible when setting the flow rate of the syringe pump. In the first try, the subjects accomplished the task within 7 seconds (IQR 6-7) and 12 seconds (IQR 3-14; *P*=.60) using the ANWS and the conventional system, respectively. As expected, the response time decreased significantly in further repetitions, especially for the conventional setting; the median was 4 seconds (IQR 3-6.75) for the conventional setting and 4 seconds (IQR 3-4.75) for the ANWS setting (*P*=.72).

During the implantation of prosthetic sockets in total hip arthroplasties and in surgeries in general, several complications may occur. Within the surgery simulation, a pulmonary embolism was mimicked (see Table 1). The intelligent alarm, integrated in the ANWS, notified the anesthesiologist immediately about the potential occurrence of this particular complication due to changes in vital signs. During the conventional trial, 6.8 seconds (IQR 2-10) elapsed before participants detected the complication.

Eye Tracking and the Think-Aloud Method

All areas of interest were identified by each participant. However, the usability study demonstrated that improvements in some elements of the GUI were still necessary. When starting the surgery workflow, the button used to select the patient and the corresponding operation, in this case a total hip arthroplasty, showed a clear case of hidden affordance. Furthermore, it was not intuitively obvious which button needed to be clicked to update the workflow process. Although eye movements indicated searching and the participants often looked directly at the correct button, they had difficulties in understanding its purpose. Another significant case of hidden affordance was recognized when dragging and dropping the device panels from the device selector (see Figure 1b) to the central region of the GUI (see Figure 1d). Otherwise, no “vampire effects” were detected.

Questionnaire: Part B

Figure 5 shows the evaluation of a segment of the questionnaire. In general, a good acceptance of the ANWS functional model was observed. In fact, in the SUS, all features were graded as at least *excellent* as follows: decision-support system, 95 points (IQR 87.5-97.5); workflow management, 97.5 points (IQR 82.5-97.5); SSC, 100 points (IQR 87.5-100); cross-device interaction, 97.5 points (IQR 92.5-100); and unified surface, 100 points (IQR 92.5-100).

Figure 4. Tukey’s box plots comparing the response times of both groups: anesthesia workstation (ANWS; left) versus conventional setting (right). The “o” represents for outliers, “x” marks the arithmetic mean, and “*” indicates statistical significance (*P*<.05) between the ANWS and conventional groups. To analyze the learning curve, response times for the first try and further repetitions were compared. (a) Time required to read out medical parameters. (b) Time required to set the positive end-expiratory pressure. (c) Time required to set the flow rate of the syringe pump.

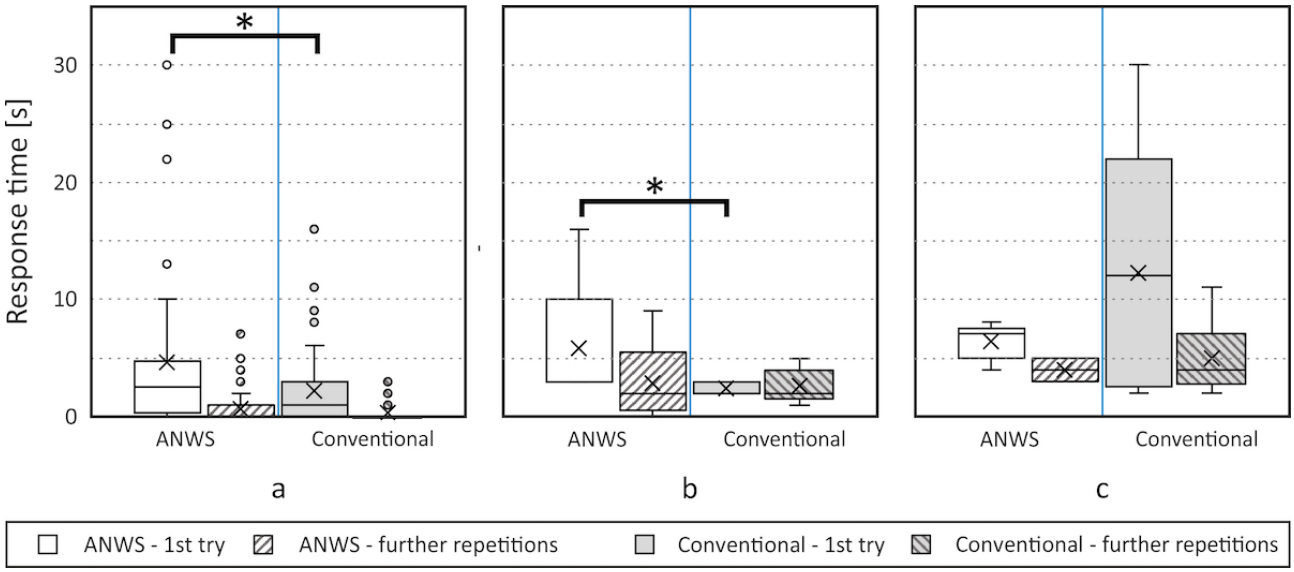
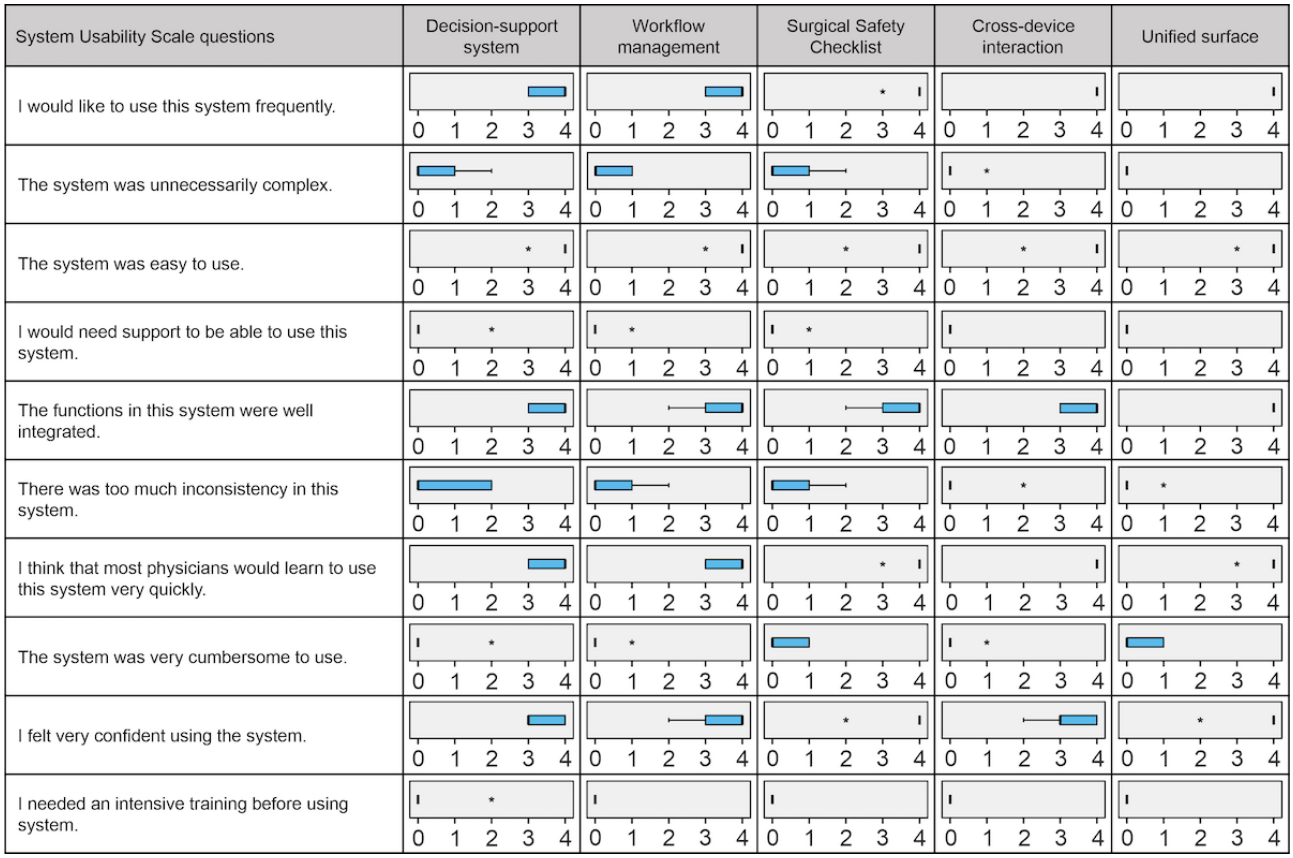


Figure 5. The second part of the questionnaire: the System Usability Scale results presented by Tukey's box plots. Each feature was rated from 0 (*strongly disagree*) to 4 (*strongly agree*). Asterisk indicates an outlier.



In the questionnaire, the volunteers were also asked to sort ANWS features—tested and untested—according to their preference, from 1 (*favorite*) to 5 (*least favorite*). Figure 6 (top and bottom) shows the results for tested and untested features, respectively. Regarding the tested features, no unanimity was observed (ie, the preference varied between the subjects). Although the subjects did not test the feature *automatic documentation*, the great majority found it meaningful (see Figure 6). *Compilation of all alarms* and *control devices from other departments* were the second-favorite features of the anesthesiologists.

At the end of the questionnaire, the subjects were asked to choose, for each criterion, their favorite system: ANWS or conventional. The results, represented in Figure 7, demonstrated that they favored the ANWS functional model. In their opinion, it would permit them to do the following:

1. Better monitor the parameters of different devices.
2. Always complete the SSC.
3. Get important additional information.
4. Improve the overall usability of the connected devices.
5. Get a better overview of the current surgical step.
6. Increase patient safety.

In general, the subjects demonstrated clear opinions with the exception of subject 2; he had a neutral opinion with regard to item 4. Although the majority of the subjects considered the ANWS functional model as their personal favorite system, 2 out of 5 (40%) considered the use of the ANWS to be more complicated when rating the *monitoring of parameters from different devices is easier* criterion. Furthermore, during an emergency surgery, 4 out of 5 doctors (80%) would prefer to use the conventional system.

Figure 6. Tested (top) and untested (bottom) features were rated by the subjects from 1 (*favorite*, dark blue) to 5 (*least favorite*, dark gray).

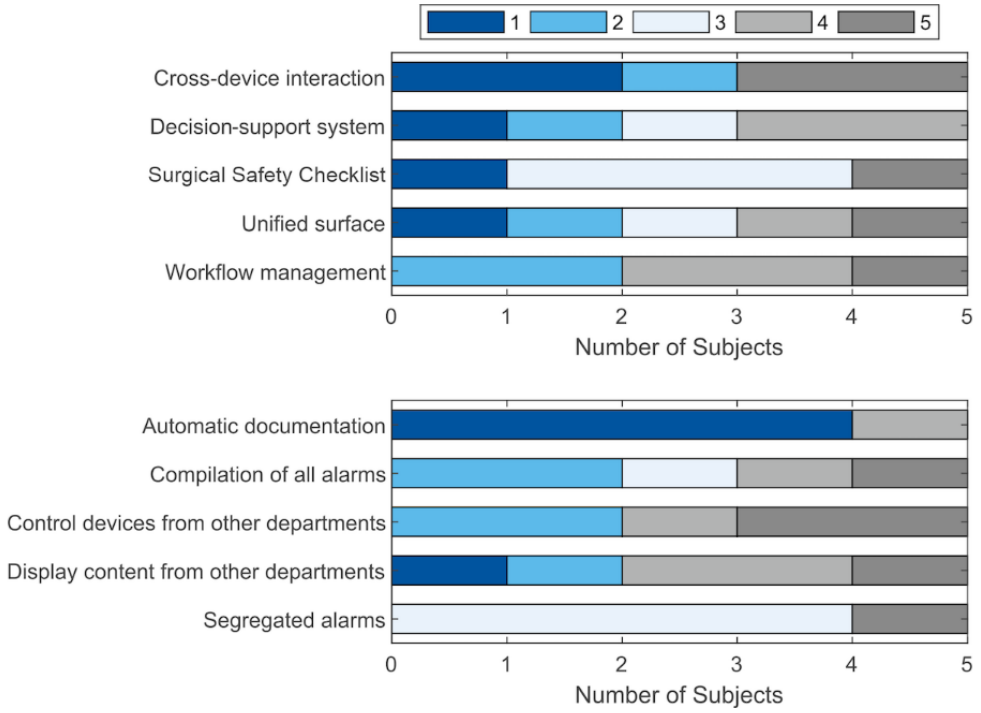
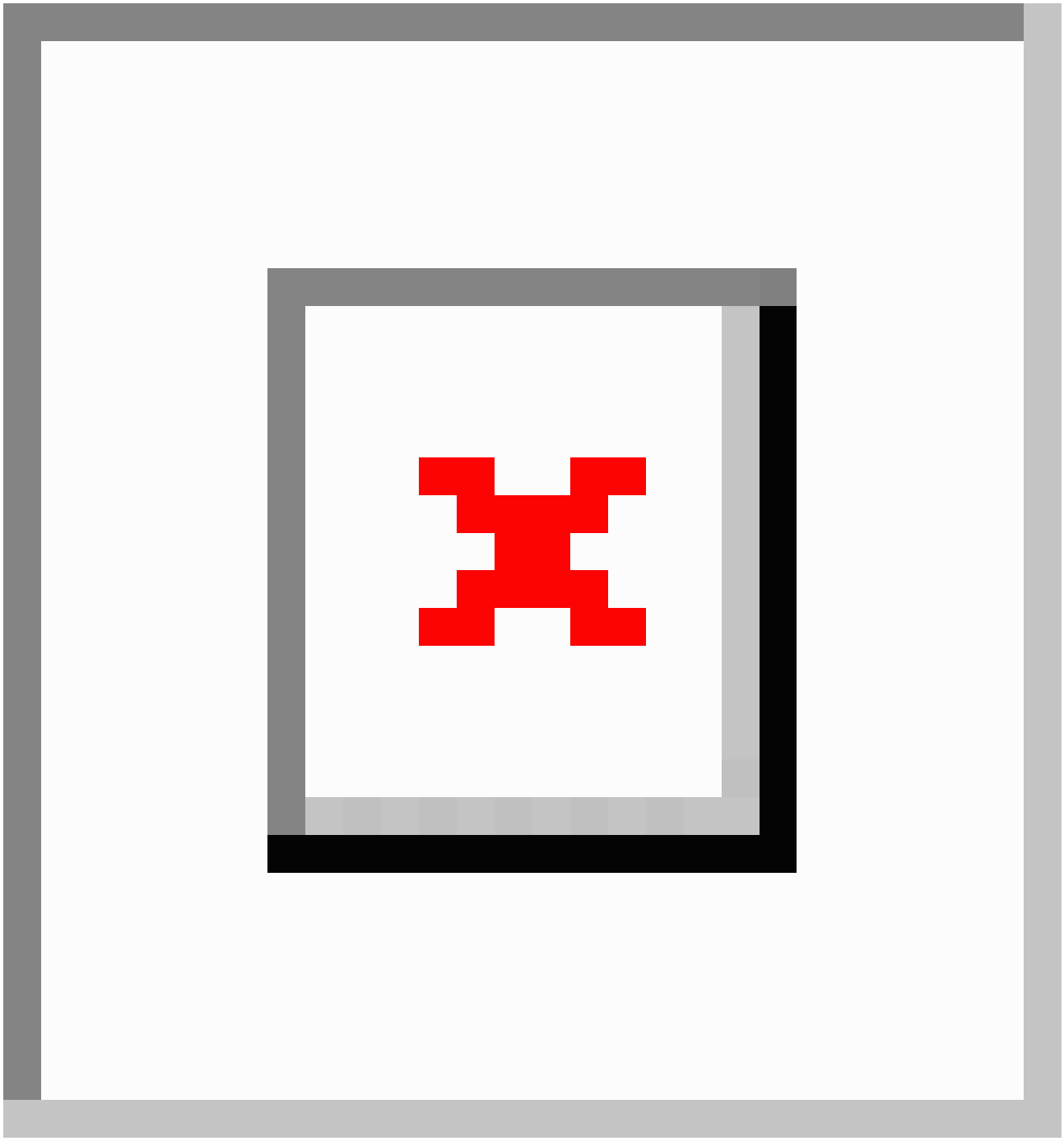


Figure 7. Graphical representation of the subjects' favorite system, anesthesia workstation or conventional systems, according to seven criteria.



Discussion

Principal Findings

In this study, 5 anesthesiologists were requested to test and evaluate a newly self-developed ANWS functional model. To test its usability, the subjects were requested to perform several tasks using the ANWS and the conventional system. The ANWS was deemed noninferior compared to the well-known system that was used daily.

All subjects considered themselves *enthusiastic about new technologies* and their technical knowledge ranged between *normal* and *proficient*. This was confirmed during the study. In general, they did not demonstrate having difficulties while solving the tasks with the ANWS functional model and accomplished them successfully.

To analyze the learning curve, some tasks were repeated. As expected, during the first try, the subjects requested more time (2.5 seconds [IQR 0.75-4.25] vs 1 second [IQR 0-3]) when using the ANWS. During further repetitions, the response time decreased significantly, resulting in a time span of less than 1 second in each setting. As demonstrated in Figure 4, the response times during further repetitions were very similar for the conventional and ANWS settings. These results suggest that, before using the new functional model, initial training is necessary. However, they also show a steep learning curve for physicians with normal-to-proficient technical knowledge. In this study, we also observed that this learning curve is similar

for other medical devices. One of the tasks consisted of manually setting the flow rate of a syringe pump. Since some of the participants were not familiar with operating this specific device, greater response times were observed in the first try. Once more, this emphasizes that for any medical device, initial training should be mandatory.

Unfortunately, hospitals frequently utilize medical devices from various manufacturers, even for the same appliance classes, such as patient monitoring, ventilators, and syringe pumps, which means different modus operandi. Consequently, the physician must be capable of working with all of them and of changing over without any time delay. As a result, *information overload* commonly leads to user errors; this can be ameliorated by medical device interoperability. Toward this end, a unified GUI would be especially beneficial in order to reduce (1) the number of training sessions, (2) response times, (3) potential use errors, and (4) medical costs.

Figure 4a contains a few outliers for the ANWS and the conventional system. A data point is considered an outlier if it exceeds or is 1.5 times the IQR above the 3rd quartile or 1.5 times the IQR below the 1st quartile. Due to the relatively low number of participants, outliers are expected; even so, no significant difference between the conventional and ANWS settings occurred. In each system, 40 data points for the first try and 70 data points for further repetitions were collected.

Another important advantage of the ANWS is the fact that the World Health Organization SSC is integrated into the surgical workflow panel so that it must be filled in. Therefore, this feature contributes to improved patient safety and quality of care since SSCs are often neglected in clinical practice. In addition, the functional model also includes a decision-support system aimed at assisting the medical team in the decision-making process. It uses patient data (eg, vital parameters) as input measures and combines them with mathematical models and algorithms. When potential abnormal changes are detected, the physicians are informed (eg, by an alarm). As demonstrated in the Response Time section of the Results, the anesthesiologists required more time to detect the pulmonary embolism when using the conventional system. In contrast, when using the ANWS, the subjects were promptly informed about this complication and could start with the treatment directly. Of course, the decision-support system does not replace the physician, but is able to support him or her and, therefore, improve the quality of care.

Regarding the questionnaire, the subjects considered automatic documentation a very important feature since it contributes to a reduction in the workload. Concerning the remaining features, no consensus was reached because preferences are usually subjective and vary from individual to individual. However, in general, a great acceptance of the ANWS functional model was observed. Indeed, all tested features were graded as at least *excellent*. The majority of subjects agreed that the ANWS might permit the monitoring of parameters of different devices in a more effective and efficient way and allow physicians to do the following: always complete the SSC, get important additional information, improve the overall usability and safety within the usage of the connected devices, get a better overview of the current surgical step, and increase patient safety. Based on these reasons, they chose the functional model as their personal favorite. Despite that, 4 out of 5 subjects (80%) would still prefer to use the conventional system. This is reasonable, since the ANWS is neither ready for market nor clinically approved. In addition, more meticulous training would be necessary. Conversely, 1 subject (20%) stated they would use the ANWS straight away. This unexpected answer probably indicates a halo effect.

Despite the positive feedback, there is still room for improvement. Taking the eye-tracking and think-aloud method analyses into account, two controls indicated hidden affordance: the workflow management controls and the drag-and-drop mechanism of the device panels need to be adapted to improve usability. Fortunately, no “vampire effects” were detected and all areas of interest were identified clearly.

Limitations

This study has certain limitations, which need to be addressed. First, due to the early stage of development of the ANWS, this pilot study was carried out with only 5 participants, which is the minimum requirement for formative usability tests within the usability engineering process for medical devices, according to the International Electrotechnical Commission (IEC) 62366 standard. Therefore, the statistical capacity is still limited for this pilot study.

Second, not all features of the functional model were investigated (eg, automatic documentation). Third, the experimental settings were not very close to reality, especially when considering the ANWS testing. Due to the lack of SDC-compatible devices, all measures were simulated digitally but were not derived from a patient or a patient simulator. Finally, the functional model is not certified as medical device, hence, a simulator instead of *real human subjects* was used to provide the data for the ANWS.

Despite these drawbacks, this study is, to the best of our knowledge, the first analysis regarding the acceptance of an innovative and pioneering ANWS, based on a manufacturer-independent communication protocol, namely SDC.

Although the ANWS only turned out to be noninferior compared to the conventional setting, this is already promising. We believe that the ANWS would have been superior if the following had been provided:

1. Adequate training of the professional users, analogous to the conventional system.
2. Optimization of the GUI in terms of usability aspects and design.
3. Integration into a clinical environment including *real* patient-derived measures.
4. Actual tasks during real anesthesiologic workflows, such as anesthesia induction, maintenance, and complications.
5. Complex scenarios that benefit from *smart alarms* and decision-support engines.

Conclusions

Although the technologizing of hospitals, especially in operation theaters, is increasing, medical devices and IT systems mainly work as stand-alone versus networked systems. Consequently, progress in health care digitization is very slow. In this project, we developed an anesthesia dashboard, enabling various sophisticated features based on open device interconnection. In the meantime, our project team contributed to the development and approval of SDC, a manufacturer-independent IEEE 11073 standard for medical device networking, which was approved as a worldwide accepted standard in January 2019.

The way is now paved for manufacturers to equip their medical devices with SDC-compliant interfaces, enabling interconnectivity in the operation theater and elsewhere in health care.

Acknowledgments

The project was supported by the German Federal Ministry of Education and Research. In addition, we would like to thank Uwe Wainwright and Michael Schauwinhold for their support in operating the AIXTRA Patient Simulator.

Conflicts of Interest

MO and MC are employed by Docs in Clouds GmbH (Aachen, Germany), providing telemedical services and consulting regarding smart hospital solutions. MC is CEO of Ilara GmbH (Herzogenrath, Germany), a provider of medical software, including OR.NET device simulators and connectors. AJ and MC are members of the executive board of the nonprofit association OR.NET eV.

References

1. Checketts MR, Alladi R, Ferguson K, Gemmell L, Handy JM, Klein AA, Association of Anaesthetists of Great Britain and Ireland. Recommendations for standards of monitoring during anaesthesia and recovery 2015: Association of Anaesthetists of Great Britain and Ireland. Anaesthesia 2016 Jan;71(1):85-93 [FREE Full text] [doi: 10.1111/anae.13316] [Medline: 26582586]
2. Imhoff M, Fried R. The crying wolf: Still crying? Anesth Analg 2009 May;108(5):1382-1383. [doi: 10.1213/ane.0b013e31819ed484] [Medline: 19372311]
3. Kasparick M, Golatowski F, Timmermann D. A safe and interoperable distributed alarm notification system for PoC medical devices using IEEE 11073 SDC. In: Proceedings of the 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT). 2017 Presented at: 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT); November 6-8, 2017; Bethesda, MD p. 257-261. [doi: 10.1109/HIC.2017.8227633]
4. Faiola A, Srinivas P, Duke J. Supporting clinical cognition: A human-centered approach to a novel ICU information visualization dashboard. AMIA Annu Symp Proc 2015;2015:560-569 [FREE Full text] [Medline: 26958190]
5. Koeny M, Benzko J, Czaplik M, Walter M, Radermacher K, Rossaint R, et al. Getting anesthesia online: The smartOR network. Int J Adv Internet Technol 2012;5(3 and 4):114-125 [FREE Full text]
6. Koeny M, Leonhardt S, Czaplik M, Rossaint R. On the road to predictive smart alarms based on a networked operating room. In: Proceedings of the 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS). 2012 Presented at: 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS); June 20-22, 2012; Rome, Italy p. 1-4. [doi: 10.1109/CBMS.2012.6266381]
7. Koeny M, Czaplik M, Walter M, Rossaint R, Leonhardt S. A new telesupervision system integrated in an intelligent networked operating room. In: Proceedings of EMERGING 2011, The Third International Conference on Emerging Network Intelligence. 2011 Presented at: EMERGING 2011, The Third International Conference on Emerging Network Intelligence; November 20-25, 2011; Lisbon, Portugal p. 39-44 URL: https://thinkmind.org/download.php?articleid=emerging_2011_2_50_40126
8. Balust J, Macario A. Can anesthesia information management systems improve quality in the surgical suite? Curr Opin Anaesthesiol 2009 Apr;22(2):215-222. [doi: 10.1097/ACO.0b013e328324b9e6] [Medline: 19390248]
9. Shieh JS, Linkens DA, Peacock JE. Hierarchical rule-based and self-organizing fuzzy logic control for depth of anaesthesia. IEEE Trans Syst Man Cybern C Appl Rev 1999 Feb;29(1):98-109. [doi: 10.1109/5326.740673]
10. Fan SZ, Yeh JR, Chen BC, Shieh JS. Comparison of EEG approximate entropy and complexity measures of depth of anaesthesia during inhalational general anaesthesia. J Med Biol Eng 2011;31(5):359-366 [FREE Full text] [doi: 10.5405/jmbe.820]
11. Birkle M, Benzko J, Shevchenko N. Das Projekt OR.NET: Sichere dynamische vernetzung in OP und klinik [The OR.NET Project: Secure dynamic networking in the OR and clinic]. Dtsch Z Klin Forsch, Innov Praxis 2012;11/12(6):41-45.
12. Kasparick M, Schlichting S, Golatowski F, Timmermann D. Medical DPWS: New IEEE 11073 standard for safe and interoperable medical device communication. In: Proceedings of the 2015 IEEE Conference on Standards for Communications and Networking (CSCN). 2015 Presented at: 2015 IEEE Conference on Standards for Communications and Networking (CSCN); October 28-30, 2015; Tokyo, Japan p. 212-217. [doi: 10.1109/CSCN.2015.7390446]

13. Mildner A, Janß A, Dell'Anna-Pudlik J, Merz P, Leucker M, Radermacher K. Device- and service profiles for integrated or systems based on open standards. Curr Dir Biomed Eng 2015 Sep 12;1(1):538-542 [FREE Full text] [doi: 10.1515/cdbme-2015-0128]
14. Kasparick M, Schmitz M, Andersen B, Rockstroh M, Franke S, Schlichting S, et al. OR.NET: A service-oriented architecture for safe and dynamic medical device interoperability. Biomed Tech (Berl) 2018 Feb 23;63(1):11-30. [doi: 10.1515/bmt-2017-0020] [Medline: 29346114]
15. Kasparick M, Schmitz M, Golatowski F, Timmermann D. Dynamic remote control through service orchestration of point-of-care and surgical devices based on IEEE 11073 SDC. In: Proceedings of the IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT). 2016 Presented at: IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT); November 9-11, 2016; Cancun, Mexico p. 121-125. [doi: 10.1109/HIC.2016.7797712]
16. Andersen B, Kasparick M, Ulrich H, Franke S, Schlamelcher J, Rockstroh M, et al. Connecting the clinical IT infrastructure to a service-oriented architecture of medical devices. Biomed Tech (Berl) 2018 Feb 23;63(1):57-68. [doi: 10.1515/bmt-2017-0021] [Medline: 29272252]
17. Czaplik M, Voigt V, Kenngott H, Clusmann H, Hoffmann R, Will A, Further members of the Medical Board, BMBF Research Project "OR.NET – Secure Dynamic Networking in the Operating Room and Clinic". Why OR.NET? Requirements and perspectives from a medical user's, clinical operator's and device manufacturer's points of view. Biomed Tech (Berl) 2018 Feb 23;63(1):5-10. [doi: 10.1515/bmt-2017-0043] [Medline: 29286909]
18. Duchowski AT. Eye Tracking Methodology: Theory and Practice. 2nd edition. London, UK: Springer International Publishing; 2007.
19. Holzinger A. Usability engineering methods for software developers. Commun ACM 2005 Jan 01;48(1):71-74. [doi: 10.1145/1039539.1039541]
20. Kuvita T, Karlíček M. The risk of vampire effect in advertisements using celebrity endorsement. Cent Eur Bus Rev 2014 Sep;3(3):16-22 [FREE Full text] [doi: 10.18267/j.cebr.89]
21. Jordan PW, Thomas B, Weerdmeester BA, Clelland IL, editors. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996.
22. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123 [FREE Full text]
23. Krüger A, Gillmann B, Hardt C, Döring R, Beckers SK, Rossaint R. Teaching non-technical skills for critical incidents: Crisis resource management training for medical students [Article in German]. Anaesthesist 2009 Jun;58(6):582-588. [doi: 10.1007/s00101-009-1511-6] [Medline: 19189061]
24. Nisbett RE, Wilson TD. The halo effect: Evidence for unconscious alteration of judgments. J Pers Soc Psychol 1977;35(4):250-256. [doi: 10.1037/0022-3514.35.4.250]
25. Skorning M, Bergrath S, Rörtgen D, Brokmann JC, Beckers SK, Protogerakis M, et al. E-health in emergency medicine: The research project Med-on-@ix [Article in German]. Anaesthesist 2009 Mar;58(3):285-292. [doi: 10.1007/s00101-008-1502-z] [Medline: 19221700]
26. Felzen M, Brokmann JC, Beckers SK, Czaplik M, Hirsch F, Tamm M, et al. Improved technical performance of a multifunctional prehospital telemedicine system between the research phase and the routine use phase: An observational study. J Telemed Telecare 2017 Apr;23(3):402-409. [doi: 10.1177/1357633X16644115] [Medline: 27080747]

Abbreviations

ANWS: anesthesia workstation
GUI: graphical user interface
HCI: human-computer interaction
IEC: International Electrotechnical Commission
IEEE: Institute of Electrical and Electronics Engineers
IQR: interquartile range
OR.NET: Secure Dynamic Networking in the Operating Room and Clinic
OSCLib C#: OpenStackClient C# library
PEEP: positive end-expiratory pressure
SDC: service-oriented device connectivity
SOP: standard operating procedure
SSC: Surgical Safety Checklist
SUS: System Usability Scale

Edited by G Eysenbach; submitted 19.10.18; peer-reviewed by S Zenker, M Kasparick; comments to author 27.12.18; revised version received 20.02.19; accepted 25.03.19; published 01.05.19.

Please cite as:
Ohligs M, Pereira C, Voigt V, Koeny M, Janß A, Rossaint R, Czaplik M
Evaluation of an Anesthesia Dashboard Functional Model Based on a Manufacturer-Independent Communication Standard: Comparative Feasibility Study
JMIR Hum Factors 2019;6(2):e12553
URL: http://humanfactors.jmir.org/2019/2/e12553/
doi:10.2196/12553
PMID:31042150

Original Paper

Exploring the Experiences of Individuals Allocated to a Control Setting: Findings From a Mobile Health Smoking Cessation Trial

Ulrika Müssener[1], PhD; Catharina Linderoth[1], BSc; Marcus Bendtsen[1], PhD

Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Ulrika Müssener, PhD
Department of Medical and Health Sciences
Linköping University
Building 511-001, Entrance 75
Linköping, S-58183
Sweden
Phone: 46 732702426
Email: ulrika.mussener@liu.se

Abstract

Background: Tobacco smoking is the primary cause of preventable premature disease and death worldwide. Evidence of the efficacy of text messaging interventions to reduce smoking behavior is well established, but there is still a need for studies targeting young people, especially because young adult smokers are less likely to seek treatment than older adults. A mobile health intervention, Nicotine Exit (NEXit), targeting smoking among university students was developed to support university students to quit smoking. Short-term effectiveness was measured through a randomized controlled trial, which found that immediately after the 12-week intervention, 26% of smokers in the intervention group had prolonged abstinence compared with 15% in the control group.

Objective: The objective of this study was to explore the experience of being allocated to the control group in the NEXit smoking cessation intervention.

Methods: We asked students who were allocated to the control group in the main NEXit randomized controlled trial to report their experiences. An email was sent to the participants with an electronic link to a short questionnaire. We assessed the distribution of the responses to the questionnaire by descriptive analysis. We analyzed free-text comments to 4 questions.

Results: The response rate for the questionnaire was 33.8% (258/763), and we collected 143 free-text comments. Of the responders, 60.9% (157/258) experienced frustration, disappointment, and irritation about being allocated to the control group; they felt they were being denied support by having to wait for the intervention. Monthly text messages during the waiting period thanking them for taking part in the trial were perceived as negative by 72.3% (189/258), but for some the messages served as a reminder about the decision to quit smoking. Of the responders, 61.2% (158/258) chose to wait to quit smoking until they had access to the intervention, and 29.8% (77/258) decided to try to quit smoking without support. Of the respondents, 77.5% (200/258) claimed they were still smoking and had signed up or were thinking about signing up for the smoking cessation program at the time of the questionnaire.

Conclusions: Most of the respondents reported negative feelings about having to wait for the support of the intervention and that they had decided to continue smoking. A similar number decided to wait to quit smoking until they had access to the intervention, and these respondents reported a high interest in the intervention. Free-text comments indicated that some control group participants believed that they had been excluded from the trial, while others were confused when asked to sign up for the intervention again.

Trial Registration: ISRCTN Registry ISRCTN75766527; http://www.isrctn.com/ISRCTN75766527 (Archived by WebCite at http://www.webcitation.org/7678sUKbR)

(JMIR Hum Factors 2019;6(2):e12139) doi:10.2196/12139

KEYWORDS

tobacco smoking; smoking cessation; students; text messaging; mobile phones; cell phone; control groups

Introduction

Background

Tobacco smoking is the primary cause of preventable premature disease and death worldwide. Tobacco use is estimated to kill 7 million people per year. If current trends continue, by 2030 tobacco will contribute to the deaths of more than 8 million people a year, with 80% of those deaths predicted to occur in the developing world [1]. Most smokers start in their teens and, as tobacco use increases with age, the earlier a person starts smoking, the higher their risk of becoming addicted and developing illnesses due to smoking [2]. Identifying effective interventions to help young people to quit smoking would have a major impact on population health, in both the short and the long term.

A growing body of evidence has accumulated in support of the efficacy of short message service (SMS) text messaging programs on mobile phones for health behavior change, including smoking cessation [3-6]. A Cochrane review of 12 such studies concluded that mobile interventions doubled the chances of long-term quitting compared with control groups [4]. A metareview of 13 studies on text message interventions for smoking cessation showed that cessation rates for the intervention group were 36% higher than for the control group [7]. An earlier Cochrane review of 28 trials suggested that interventions that have been shown to be effective among adults, such as motivational enhancement, could also be effective among adolescents. However, the review also found that there is insufficient evidence to recommend a specific method of intervention for young people and that more data are needed on long-term cessation [8].

Thus, the evidence for the efficacy of text messaging interventions to reduce smoking behavior is well established, but there is still a need for studies targeting young people, especially because young adult smokers are less likely to seek treatment than older adults [9]. In addition, most of the policies and smoking programs implemented in schools and universities are preventive, and thus are effective in reducing the initiation and prevalence of smoking among adolescents, rather than supporting young smokers to stop smoking [10].

Objective

In our previous research, we have developed a mobile health intervention, Nicotine Exit (NEXit), targeting smoking among university students (ISRCTN75766527) [11,12]. We previously tested the effectiveness of the intervention in a randomized controlled trial (RCT) and reported our results [13], as well as the satisfaction and acceptability of the intervention [12]. In RCTs, the estimated effect of the intervention under study can only be understood relative to the control setting. An effect size is estimated comparing 2 groups, 1 randomly allocated to be given access to an intervention, and 1 randomly allocated to a control setting. It is natural that the intervention setting is given much focus and is often explained in detail in research papers; however, the information given to, and experiences of, the control group plays a crucial role when interpreting results. The aim of this study was to explore participants' experience of being allocated to the control group in the NEXit smoking cessation intervention.

Methods

Description of the NEXit Intervention

The NEXit intervention was developed in several steps based on existing recommended smoking cessation manuals used in Sweden, previous research, and a taxonomy on behavior change techniques developed by Michie et al [14].

Elements included in the intervention were as follows: making a public declaration about quitting (ie, telling friends about the quit attempt), encouraging asking for support from family and friends, distraction techniques, and the possibility of requesting more text messages when the participant experienced strong cravings or a temporary relapse. The core program lasted for 12 weeks and consisted of 157 messages [13].

Ethics Approval

The study was approved by the Regional Ethical Committee in Linköping, Sweden (Dnr 2014/217-31).

Study Population and Procedure

To get information for subsequent revisions and improvements from the individual user's perspective, we invited college and university students participating in the main NEXit RCT to give feedback after completing the 12-week intervention and after participating in the formal follow-up of the RCT [13]. We recruited the participants from all colleges and universities in Sweden except one university that participated in a pilot study. The participants came from all levels and disciplines. Recruitment of participants was completed over a 3-week period (October 23 to November 13, 2014). A total of 827 participants were allocated to the intervention group and 763 to the control group (a waiting list group that were asked to quit on their own, but were also told that they were going to be given access to the intervention after the trial). Data on the primary outcome were collected over a 4-week period from 94% of the control group. Reminders were sent over the same 4-week period. Nonresponders to the follow-up were sent up to 6 reminders by email and 3 reminders by SMS text messaging, and 10 attempts were made to telephone those who had still not responded.

After the follow-up procedure, the control group received an email with an electronic link to a short questionnaire, with 2 reminders sent 1 week apart to nonresponders.

Questionnaire

We asked the participants in the NEXit control group 4 questions, with 3 to 4 fixed response options. The questions were about experiences of being randomly allocated to the control group, smoking behavior when not given access to the intervention, and willingness to sign up after the trial. A free-text option after each question gave the participants an opportunity to describe other factors of importance not covered by the fixed response options.

We explored experiences of being randomly allocated to a control group by 2 questions: (1) experience of having to wait for the intervention (response options: frustrating, irritating, or

disappointing because I was prepared to quit smoking; positive, inspiring, or motivating because I was given a chance to reflect on my reasons to quit smoking; it did not matter; do not know); (2) perception of receiving monthly SMS text messages during the waiting time thanking them for participating in the trial (response options: good, they made me feel part of the study and reminded me that I would get support later; bad, because I did not have access to the program; did not matter to me, I did not care about the messages; do not know).

We explored actions taken when being allocated to the control group by 1 question: (3) reaction when not given access to the intervention immediately (response options: I decided to try to quit smoking without help; I decided to postpone my quit attempt until I had access to the intervention; I used other support).

We explored willingness to sign up after the trial by 1 question: (4) intention to sign up after the trial (response options: yes, I still smoke and I have signed up for the intervention; yes, I still smoke and I am still thinking about signing up; no, I still smoke but do not want any help; no, I have quit smoking).

Statistical Analysis

We performed descriptive analysis of the distribution of the responses to the 4 questions in 4 steps. In the first step, all free-text comments to each question were read by the first and second authors (UM and CL). In the second step, CL chose a variety of the most crucial free-text comments for each question. In the third step, UM verified the chosen free-text comments and added some comments relevant to the aim of the study. In the fourth step, all authors discussed all of the chosen free-text comments, from which we selected comments that captured the main content of the specific question with regard to the aim of the study.

Results

Participants' Characteristics and Response Rate

The baseline characteristics of the participants were similar to those of nonparticipants concerning sex, age, and marital status (Table 1). Further, the baseline characteristics of the control group participants were similar to those of the intervention group in the main study [13] concerning sex, age, and marital status. Thus, we regarded the participants in this study as being broadly representative of the intervention group in the RCT. The overall response rate was 33.8% (258/763).

About a third (83/258, 32.2%) of the respondents provided 143 comments to the 4 questions; the other 67.8% (175/258) did not offer any additional comments. Most comments were on the question regarding how participants reacted when they did not get access to the intervention at once (42/143, 29.4%). The fewest comments (n=26/143, 18.2%) were provided for the question on willingness to sign up for the intervention after the trial. On average, approximately 35 comments were provided for each question.

We report the responses to the 4 questions, as well as citations from the free-text comments to illustrate and underline the pattern of responses to the fixed response options. The comments were translated from Swedish into English by the first author. The designation after each quotation is the code assigned to the participant.

Experiences of Having to Wait for the Smoking Cessation Aid

Of the respondents, 60.9% (157/258) experienced frustration, disappointment, and irritation when being told that they had to wait for the novel smoking cessation aid. In the free-text comments, some participants emphasized that being allocated to the control group and thus having to wait 4 to 5 months for access evoked feelings of being denied support and decreased their motivation to stop smoking.

> *I felt quite dejected; to first make the decision to participate in the program, but then being denied.* [Participant 212]

> *It felt like taking a step back, and it had a demotivating effect on my choice to quit smoking.* [Participant 210]

However, 10.9% (28/258) reported that having to wait for support was positive and inspiring because they had the chance to reflect on their decision to quit smoking or had time to really set a goal to quit. In their comments, some described that having to wait for smoking cessation increased their chances of succeeding and that their motivation increased.

> *I was prepared to quit immediately, but due to the delay my subconscious has set a goal for my smoking cessation, which increases my chances.* [Participant 149]

> *Initially I felt disappointed, but then I thought that this will turn out just fine because I will get the support later; meanwhile my decision to quit smoking has really engrained itself. It feels like my motivation to quit is much more rooted than it ever has been before.* [Participant 156]

Of the respondents, 28.3% (73/258) stated that having to wait for the support did not matter or they did not know whether it mattered.

Perceptions of Receiving Text Messages During the Waiting Time

All participants allocated to the control group received monthly text messages during the waiting time thanking them for taking part in the trial. Of the respondents, 72.3% (189/258) thought these messages were bad or worthless. Among those, some highlighted that the text messages were confusing because these participants did not have access to the intervention but were still being thanked for taking part.

Table 1. Baseline characteristics of responders and nonresponders (N=763).

Characteristics	Nonresponders (n=505), n (%)	Responders (n=258), n (%)	*P* value[a]
Female	332 (65.7)	190 (73.6)	.03
Age (years)			.31
<21	56 (11.1)	18 (7.0)	
21-25	209 (41.4)	117 (45.3)	
26-30	116 (23.0)	60 (23.3)	
≥31	124 (24.6)	63 (24.4)	
Single	309 (61.2)	150 (58.1)	.46

[a]Pearson chi-square test with Yates continuity correction.

> *It felt like I was being reminded that I could have been part of the study, but was not allowed.* [Participant 212]
>
> *It was mostly annoying to get a bunch of SMS saying that I was part of the study, which I wasn't!* [Participant 134]
>
> *I did not understand what the purpose of the messages was. Was I part of the study and had missed something?* [Participant 58]

A total of 22.5% (58/258) of the respondents stated that these text messages contributed to feelings of participation and served as a reminder of receiving the smoking cessation intervention later. Some participants expressed that just receiving text messages reminding them of the intervention was helpful. But others claimed that there was a risk of becoming tired of waiting.

> *It was a reminder of smoking...and I started to feel strongly about quitting.* [Participant 114]
>
> *It was very good to be reminded, but when you want help you get tired of waiting.* [Participant 77]
>
> *Regardless of upcoming support, it affected my perception of smoking negatively. I therefore smoked less.* [Participant 55]

Of the respondents, 31.4% (81/258) stated that the text messages did not matter or that they did not care about the messages.

Actions Taken When Being Allocated to the Control Group

More than half of the respondents in the control group (158/258, 61.2%) chose to wait to quit smoking until they had access to the intervention and, of those, 72.8% (115/158) experienced receiving text messages during the waiting time as negative.

Among those who chose to continue to smoke, some claimed that they had wanted to quit smoking but that they needed support, and not being given access to the program was a reason to wait.

> *I feel like I want to quit, but I can't put my back into it! The text messages would have been a push in the right direction.* [Participant 9]

Many comments were about trying to stop or reducing smoking after being randomly allocated to the control group, and 29.8% (77/258) of respondents decided to try to quit smoking without support. Some succeeded in quitting smoking and some cut down but had a relapse after a period without cigarettes.

> *I cut down considerably, even completely stopped for about 1 month.* [Participant 121]
>
> *I didn't smoke for 4 weeks, then at a party I started again.* [Participant 226]
>
> *I tried but unfortunately it did not work, now I smoke again.* [Participant 196]

Only 8.9% (23/258) of respondents reported that they used other support during the waiting time, mostly nicotine aids.

Willingness to Sign Up for Smoking Cessation After the Trial

Of the respondents, 77.5% (200/258) claimed they were still smoking and had signed up, or were thinking about signing up, to the smoking cessation program. The need for support was expressed in different ways, and the reasons for signing up after the trial included not feeling well, having a disease, and needing support to quit.

> *Please help me, I feel unwell and really want to quit smoking. I have a disease and my symptoms may get better if I do not smoke.* [Participant 226]
>
> *I have cut down my smoking but need help to quit completely.* [Participant 150]

Among the respondents who claimed they were thinking about signing up, some were concerned about the sign-up procedure and expressed confusion about whether they needed to sign up again or if the first sign-up when entering the program still counted.

> *...but what, sign up again? I have already signed up once?* [Participant 1657]
>
> *I thought it (the program) would start automatically when it was my turn. I already signed up for the support, but I didn't have access to it.* [Participant 218]
>
> *I didn't know I needed to sign up; wasn't it just to accept participation?* [Participant 137]

Only 8.1% (21/258) of respondents answered that they had decided not to sign up for the support although they still smoked. A total of 14.3% (37/258) responded that they had quit smoking

during the waiting time and were not interested in signing up for the support.

Discussion

In RCTs, the estimated effect of the intervention under study can only be understood relative to the control setting; thus, the aim of this study was to explore the experience of being allocated to the control group in the NEXit smoking cessation intervention.

Principal Findings

The main findings of this study are that most of the participants experienced frustration, disappointment, and irritation about being allocated to the control group. Monthly text messages during the waiting period were perceived as poor and pointless, and not being given the intervention at once was misunderstood as being excluded from the trial. Most of the respondents decided to wait to quit smoking until they had access to the intervention. There was high interest in the novel intervention after the trial.

In the NEXit trial, we decided to ask the control group to quit smoking on their own, but also to inform them that the novel intervention would be available to them after the trial was finished. While a waiting list approach reduces ethical dilemmas, it can also create a feeling of missing out or resentment in the control group. Approximately 61% of the respondents in the control group reported that they had negative feelings about having to wait for the support, and a similar proportion of respondents reported that they had decided to continue smoking while waiting for access to the intervention. While the response rate to the questionnaire was low (34%), having negative feelings about having to wait for the support is still a matter of concern that should be addressed in future studies. Previous research showed that individuals who sign up for lifestyle intervention trials have previously tried to change their behavior but were not able to do so using existing means. They then feel disappointed that they are not given access to a new support tool that they believe would help them, regardless of the fact that the intervention under trial has yet to be proven to be effective [15]. Because the control group behaves in a manner that was not initially planned for, this might create a potential bias in the effect measurement of the intervention, one that cannot be estimated using the primary outcome data collected during trials.

To alleviate the feeling among control group participants of being dismissed or left out, we decided that they were to receive messages throughout the trial to remind and thank them for being part of the trial. Nearly 73% of the respondents found these messages to be poor and pointless. Free-text comments indicated that respondents felt that they were reminded about missing out on the intervention. Furthermore, free-text comments revealed that participants perceived being allocated to the control group as not being allowed to be part of the trial. Thus, not being given the intervention was misunderstood as being excluded from the trial. The monthly reminders were supposed to alleviate the feeling of being left out but seem to have become a reminder of just that.

There was high interest in the novel intervention after the trial, with approximately 71% of responders reporting that they still smoked and had signed up for the intervention. It is, however, interesting that participants were confused as to whether signing up for the trial also meant that they would be given the intervention afterward without actively asking for it. Thus, there seems to be a dichotomy between those who thought that they were excluded and those who thought that they were waiting.

The responses to the questionnaire sent to control group participants strengthen a growing body of evidence that suggests that participants in RCTs are not well aware of the trial design [16-19]. The difficulties include explaining to participants that the intervention under trial is yet to be shown to be effective: control group participants in the NEXit trial felt like they were missing out on support, yet it was not known whether the support would be effective at the trial start. Explaining the concept of placebos and allocation to different groups also presents challenges [20,21].

We took several steps to ensure validity of the results and to prevent bias in the selection of the 143 free-text comments. Free-text comments were read by the first and second authors independently many times. Free-text comments were first selected separately, and then compared and discussed among the authors. One author verified the chosen free-text comments and added some comments relevant to the aim of the study. We excluded free-text comments not agreed on by all authors. Authors were of different ages, sexes, and educational backgrounds.

Limitations

Limitations include the low response rate: only 34% of the control group responded to the questionnaire regarding their experiences. However, most of the respondents reported negative feelings and comments about being allocated to the control group. This implies that, even if those who did not respond were different from those who did, a large part of the control group were dissatisfied with their participation.

Conclusion

Future studies should more carefully consider not only the control setting, but also how it is presented to the control group, in order to reduce friction and better reflect the control reference of interest. In the NEXit trial, it would have been advisable to present the trial to participants as comparing 2 different interventions: 1 intervention consisting of immediate access to the SMS text messaging support, and 1 intervention consisting of a motivational phase during which participants would be asked to increase their motivation and attempt to quit smoking on their own with the use of existing support tools, and then be given the SMS support. Hypothetically, this would give the control group a feeling of being part of the trial, having been allocated to an intervention, rather than being told that they were on their own. This would also make the control group more homogeneous, not creating a dichotomy between those who believe they are part of the trial and those who do not. Effect sizes should then be interpreted in light of this control setting.

Acknowledgments

The study was funded by the Swedish Research Council, grant number 521-2012-2865.

Authors' Contributions

MB designed the study. UM took a leading role in developing the intervention and UM and CL in recruiting the participants. MB did all the programming and supervised the data collection and all technical aspects of the delivery of the intervention. CL and UM wrote the first draft of the manuscript except for the discussion section that was written by all authors. UM revised the draft.

Conflicts of Interest

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

References

1. Tobacco. Geneva, Switzerland: World Health Organization; 2018 Mar 09. URL: http://www.who.int/mediacentre/factsheets/fs339/en/ [accessed 2019-02-11] [WebCite Cache ID 7676u6ZNk]
2. Alkherayf F, Wai EK, Tsai EC, Agbi C. Daily smoking and lower back pain in adult Canadians: the Canadian Community Health Survey. J Pain Res 2010 Aug 26;3:155-160 [FREE Full text] [doi: 10.2147/JPR.S11031] [Medline: 21197319]
3. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
4. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
5. Ybarra ML, Holtrop JS, Prescott TL, Rahbar MH, Strong D. Pilot RCT results of stop my smoking USA: a text messaging-based smoking cessation program for young adults. Nicotine Tob Res 2013 Aug;15(8):1388-1399 [FREE Full text] [doi: 10.1093/ntr/nts339] [Medline: 23348969]
6. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
7. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
8. Stanton A, Grimshaw G. Tobacco cessation interventions for young people. Cochrane Database Syst Rev 2013;8:CD003289. [doi: 10.1002/14651858.CD003289.pub5] [Medline: 23975659]
9. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: issues and recommendations. Nicotine Tob Res 2003 Feb;5(1):13-25. [Medline: 12745503]
10. Lamkin L, Davis B, Kamen A. Rationale for tobacco cessation interventions for youth. Prev Med 1998 Oct;27(5 Pt 3):A3-A8. [Medline: 9808812]
11. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. SMS-based smoking cessation intervention among university students: study protocol for a randomised controlled trial (NEXit trial). Trials 2015 Apr 08;16:140 [FREE Full text] [doi: 10.1186/s13063-015-0640-2] [Medline: 25872503]
12. Müssener U, Bendtsen M, McCambridge J, Bendtsen P. User satisfaction with the structure and content of the NEXit intervention, a text messaging-based smoking cessation programme. BMC Public Health 2016 Nov 22;16(1):1179 [FREE Full text] [doi: 10.1186/s12889-016-3848-5] [Medline: 27876031]
13. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):321-328. [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
14. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
15. McCambridge J, Sorhaindo A, Quirk A, Nanchahal K. Patient preferences and performance bias in a weight loss trial with a usual care arm. Patient Educ Couns 2014 May;95(2):243-247 [FREE Full text] [doi: 10.1016/j.pec.2014.01.003] [Medline: 24492159]
16. Fortun P, West J, Chalkley L, Shonde A, Hawkey C. Recall of informed consent information by healthy volunteers in clinical trials. QJM 2008 Aug;101(8):625-629. [doi: 10.1093/qjmed/hcn067] [Medline: 18487271]
17. Pesudovs K, Luscombe CK, Coster DJ. Recall from informed consent counselling for cataract surgery. J Law Med 2006 May;13(4):496-504. [Medline: 16756218]

18. Dhital R, Norman I, Whittlesea C, Murrells T, McCambridge J. The effectiveness of brief alcohol interventions delivered by community pharmacists: randomized controlled trial. Addiction 2015 Oct;110(10):1586-1594 [FREE Full text] [doi: 10.1111/add.12994] [Medline: 25988589]
19. Saitz R, Palfai TP, Cheng DM, Horton NJ, Freedner N, Dukes K, et al. Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. Ann Intern Med 2007 Feb 06;146(3):167-176. [Medline: 17283347]
20. Moodley K, Pather M, Myer L. Informed consent and participant perceptions of influenza vaccine trials in South Africa. J Med Ethics 2005 Dec;31(12):727-732 [FREE Full text] [doi: 10.1136/jme.2004.009910] [Medline: 16319239]
21. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. JAMA 2004 Oct 6;292(13):1593-1601. [doi: 10.1001/jama.292.13.1593] [Medline: 15467062]

Abbreviations

NEXit: Nicotine Exit
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 07.09.18; peer-reviewed by A Al-Zalabani, Y Albalawi, A Bedendo; comments to author 27.10.18; revised version received 05.12.18; accepted 29.12.18; published 02.04.19.

Please cite as:
Müssener U, Linderoth C, Bendtsen M
Exploring the Experiences of Individuals Allocated to a Control Setting: Findings From a Mobile Health Smoking Cessation Trial
JMIR Hum Factors 2019;6(2):e12139
URL: https://humanfactors.jmir.org/2019/2/e12139/
doi:10.2196/12139
PMID:30938682

Original Paper

Uptake and Scalability of a Peritoneal Dialysis Virtual Care Solution: Qualitative Study

Lianne Jeffs[1,2], BScN, MSc, PhD; Trevor Jamieson[3,4], MD, FRCPC; Marianne Saragosa[4], BScN, MN; Geetha Mukerji[2,4], MD, MSc, FACP, FRCPC; Arsh K Jain[5], MD, MSc, FRCPC; Rachel Man[5], MSc; Laura Desveaux[2,4], PT, PhD; James Shaw[2,4], PhD; Payal Agarwal[4,6], BASc, MD; Jennifer M Hensel[4], MD, MSc; Maria Maione[7], BScN, MN; Nike Onabajo[4], MSc, PMP; Megan Nguyen[4], MN; R Bhatia[2,4], MD, MBA

[1]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada

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

[3]St. Michael's Hospital, Toronto, ON, Canada

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

[5]London Health Sciences Centre, London, ON, Canada

[6]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[7]St. Michael's Hospital, Toronto ON, ON, Canada

Corresponding Author:
Lianne Jeffs, BScN, MSc, PhD
Lunenfeld-Tanenbaum Research Institute
Sinai Health System
600 University Avenue
Toronto, ON, M5G 1X5
Canada
Phone: 1 4165221052
Email: lianne.jeffs@sinaihealthsystem.ca

Abstract

Background: Early research in the area of virtual care solutions with peritoneal dialysis (PD) patients has focused on evaluating the outcomes and impact of these solutions. There has been less attention focused on understanding the factors influencing the uptake, usability, and scalability of virtual care for chronic kidney disease (CKD) patients receiving PD at home.

Objective: In this context, a study was undertaken to (1) assess and understand the factors influencing the uptake of a virtual care solution and (2) provide recommendations for the scalability of a virtual care solution aimed at enhancing CKD patients' outcomes and experiences.

Methods: This study used a qualitative design with semistructured interviews and a thematic analysis approach. A total of 25 stakeholders—6 patients and 3 caregivers, 6 health care providers, 2 vendors, and 8 health system decision makers—participated in this study.

Results: The following three primary mechanisms emerged to influence the usability of the virtual care solution: (1) receiving hands-on training and ongoing communication from a supportive team, (2) adapting to meet user needs and embedding them into workflow, and (3) being influenced by patient and caregiver characteristics. Further, two overarching recommendations were developed for considerations around scalability: (1) co-design locally, embed into the daily workflow, and deploy over time and (2) share the benefits and build the case.

Conclusions: Study findings can be used by key stakeholders in their future efforts to enhance the implementation, uptake, and scalability of virtual care solutions for CKD and managing PD at home.

(JMIR Hum Factors 2019;6(2):e9720) doi:10.2196/humanfactors.9720

KEYWORDS

virtual care solutions; peritoneal dialysis; qualitative research; patient-centric care; chronic kidney disease

Introduction

End-stage renal disease (ESRD) continues to prevail as a global public health problem affecting over 700,000 Americans [1] and more than 41,000 Canadians [2]. ESRD and chronic kidney disease (CKD) are frequently associated with substantial health care and societal costs [3,4]. People with CKD experience challenges accessing kidney care due to limited health care resources, which is exacerbated by geographical barriers for those patients who live in remote or underserved areas [5-8]. In this context, patients and their family members must travel long distances to obtain care, which adds stress, imposes additional costs, and contributes to the poorer quality of care [9] and, in some situations, mortality [10]. Barriers to access kidney care and associated costs may worsen as the prevalence of CKD increases with the rising aging population, who may experience multiple comorbidities that will require multiple medications to treat [4,9]. Alternate CKD care delivery models are needed to address these challenges faced by CKD patients to ensure more flexible, convenient, person-centered care delivery models, particularly for those living in rural and remote locations [11].

Peritoneal dialysis (PD) is one strategy to mitigate challenges with accessing care for managing CKD, particularly ESRD. Globally, 190,000 patients are receiving daily PD at home [12] following education and training by a specialized dialysis health care professional team [13,14]. Outcomes associated with PD include increased satisfaction [13], increased quality of life [15,16], and survival advantage, especially in the first few years of therapy [17-19], with decreased costs to the health care system [20]. Virtual care, often referred to as telehealth or telemedicine, is a rapidly evolving area where health care services are delivered, in part, through technology, including clinician-to-clinician, clinician-to-patient, and patient-to-mobile health technology communication that aims to enhance patients' self-management of their disease in a home setting [4,17,20,21]. Technology includes remote monitoring [4,21,22], mobile phones [5,23], virtual wards [24], and Web-based eHealth portals [25]. Virtual care, coupled with PD, has the potential to improve access for CKD care for patients in their own homes [26-30].

A recent systematic review reported a significant decrease in hospital readmission, emergency room visits, number of days in the hospital, and increased compliance with home self-management programs associated with the use of virtual care solutions in the CKD population [4]. To date, research in this area has focused on evaluating the outcomes and impact of these solutions. There has been less attention focused on the factors that lead to the success or failure in uptake and scalability of telehealth programs for the CKD population [21]. Gaining insight into this gap may yield useful knowledge in improving the uptake and scalability of the virtual care solution (ie, telehealth) that can be used to improve future implementation efforts and ultimately improve patient and health service utilization outcomes. In this context, as part of a larger study, the qualitative arm included an exploration of the factors influencing the uptake of, and recommendations for, scalability of a virtual care solution aimed at enhancing CKD patients' outcomes and experiences.

Methods

Study Design

This qualitative study design was employed as part of a larger evaluation of the eQ Connect (eQOL Inc) intervention that includes a parallel-arm, randomized controlled trial (RCT). The aim of the RCT, CONNECT Trial, is to determine if utilizing a virtual care solution that includes remote monitoring software, eQ Connect, improves selected clinical outcomes for PD patients. The study protocol provides more detail of the RCT [31]. Ethics approval for the qualitative study design was obtained from the Research Ethics Boards at Women's College Hospital, St Michael's Hospital, London Health Sciences Centre, and Humber River Regional Hospital, all located in Ontario, Canada.

Study Setting

Participants were recruited from two hospitals: one urban teaching hospital, London Health Sciences Centre, and one community hospital, Humber River Hospital, from Ontario, Canada. Collectively, over 200 PD patients receive care from these two centers. Eligible patients were approached during their regularly scheduled clinic visit at the PD clinics.

Theoretical Framing

The theoretical framing of the qualitative design included an integration of the following three key frameworks: (1) an adaptation of the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework [32]; (2) Institute for Healthcare Improvement's Triple Aim [33]; and (3) a scalability framework [34]. Specifically, we postulated that the successful uptake and scalability of the intervention would be influenced by a series of factors. These factors include leadership engagement and culture, communication methods, and social networks; structures including a learning system that incorporates training, support, and infrastructure that connects adopters and experts; and a data system to support measurement for improvement. Refer to the protocol for further details [31].

Virtual Care Solution Description

The virtual care solution included a remote monitoring software, eQ Connect (eQOL Inc), that provides support for patients undergoing PD. eQOL is a Canadian health technology company based in Toronto, Ontario, specializing in the development and deployment of innovative solutions to enable patients to better manage their care outside of the hospital environment. The solution provides up-to-date patient information to the health care team and aims to promote patient adherence with PD regimens. The platform consists of a patient-facing interface, Patient Portal, that operates on a tablet (ie, iPad Mini 2 running Apple iOS); the platform is designed to record and upload data (eg, treatment progress, health status, and supply usage) easily and securely over the Internet to a secure data center. Information is transferred to a compliant secure data center where clinicians can gain access to the data by logging in to the

Support Portal. A more detailed description of eQ Connect is provided in a published protocol paper [31].

Data Collection and Analysis

The qualitative component included semistructured interviews with the principal stakeholders involved in the implementation of the eQ Connect app process: patients and caregivers, health care providers, and health system decision makers. Qualitative interviews included questions about (1) participants' experiences of learning about and using the technology; (2) changes to health care provider workflow required to use the technology effectively; (3) organizational changes needed to support the technology; and (4) health system barriers and facilitators to effective implementation, evaluation, and scalability. Interviews were conducted through a telephone conversation at a time convenient to study participants between baseline and 3 months of implementation of eQ Connect from March to June 2017. The average length of time for interviews with stakeholders was as follows: patients and caregivers, 25 minutes (range 11-44 minutes); health care providers and vendors, 22 minutes (range 8-44 minutes); and health system decision makers, 49 minutes (range 44-59 minutes).

Qualitative interviews were conducted and recorded by experienced qualitative research assistants who then transcribed the interviews into Word documents, prepared the documents for qualitative analysis, and analyzed the interviews using thematic analysis [35]. Specifically, a coding schema was constructed and used to categorize the narrative text. This analytical process involved two researchers reviewing the transcripts line-by-line separately to identify sections of text that serve as codes. The researchers met to determine the codes and categories through consensus and the researchers developed themes and subthemes from the categorical data through consensus. As a final step to ensuring methodological rigor, the principal investigator reviewed all the original transcripts with the emergent coding schema. For more details on the study methods, refer to the published protocol [31].

Results

Sample Characteristics

Overall, the qualitative component involved 25 participants from the following stakeholder groups: 9 (36%) end users, including 6 (24%) patients and 3 (12%) family caregivers; 6 (24%) health care providers; 2 (8%) vendors; and 8 (32%) health system decision makers. In terms of the 9 patients and family caregivers, there were 5 (56%) males and 4 (44%) females. The average age of the patients and their caregivers was 66 years (range 41-86). Of the 9 patients and caregivers, 5 (56%) were married, 2 (22%) were common law, 1 (11%) was single, and 1 (11%) was divorced. Of the 9 patients and caregivers, 7 (78%) were educated at a university or college level and 2 (22%) at a high school level. The average length of time patients had been managing their CKD was 9.7 years (range 1-31). In terms of the 6 health care providers, 3 (50%) were project coordinators, 2 (33%) were nurses, and 1 (17%) was a physician. The vendor cohort of 2 participants included 1 (50%) product development manager and 1 (50%) clinical coordinator. No demographic information was obtained from the participating health system decision makers that drew from a variety of provincial agencies (eg, funder and networks).

Themes

The following three factors associated with the uptake of eQ Connect by patients receiving PD at home emerged: (1) receiving hands-on training and ongoing communication from a supportive team, (2) adapting to meet user needs and embedding into workflow, and (3) being influenced by patient and caregiver characteristics. Further, the following two overarching recommendations emerged for considerations around scalability of eQ Connect: (1) co-design locally, embed into the daily routine and workflow, and deploy over time and (2) share the benefits and build the case.

Influencing Factors

Receiving Hands-On Training and Ongoing Communication From a Supportive Team

This theme captured how patients and their caregivers valued the opportunity to receive a face-to-face, hands-on, brief training session on how to use the telemonitoring equipment and iPad as well as having access to, and receiving timely communication from, a supportive team (ie, health care providers in clinic and vendor). Patients described the face-to-face training session as "simple, short, and very good at orienting" them to using the tablet as part of their daily care routine. One caregiver noted the following:

> *...her teaching was good it was just a lot to take in in an hour. I was given all kinds of information the very first time the nurse talked to us and I came home and read it through and then when they called and said we were accepted [into the trial] that was fine.* [Caregiver]

Further, ongoing check-ins from the research and clinic staff ensured that the patients and caregivers, when present, knew how to use eQ Connect.

Patients also valued having access to the vendor if problems arose with the technology and being able to communicate directly with the staff at the CKD clinic, who they described as "very responsive." One patient described the following:

> *They are always there, they always call me back—I've hardly had to call them. If I have a small problem, I'll just message the support staff at the hospital and they will message me back an answer. They are always on top of it...right through the iPad.* [Patient]

Health care providers at the clinic also described the ability to connect with patients and their caregivers through the iPad, as illustrated in the following narrative:

> *It's a great way to keep connected with the patient in the home and it's also a positive thing because uh when you're looking at the iPad, it's a communication tool with the messaging center that can help address issues that patient might have without having to call us.* [Health care provider]

Adapting to Meet User Needs and Embedding Into Workflow

This theme reflects how the vendor adopted the virtual care solution to meet user—patients' and providers'—needs and embedded them into the providers' daily workflows. Patients, caregivers, and providers described how they would raise issues around technical glitches of eQ Connect and how the vendor would be receptive and adapt the functionality of the tablet to address the issues (eg, challenge entering data on the screen). This is noted in the following excerpt from a patient:

> *Whenever I did the incorrect inputting, I would be getting a phone call and they would touch base with me. I like the new way they've done the effluent screen—I told you that I kept incorrectly inputting, now it's all on one screen you can see your three different things you have to input and I do like that. I like it when it's all visual on one page. But that was certainly an improvement when they did that.* [Patient]

Another patient shared the following:

> *I gave her some ideas that I thought would make it easier and they seemed very receptive and then they [vendor] changed that.* [Patient]

The vendor also encouraged the health care team to "reach out and give feedback at any time." The vendor further noted the following:

> *I work together with the team to customize the software specifically to different sites. We meet with the nurses and doctors and find [out] about their specific needs and the demographics of their patient population.* [Vendor]

The adaptability of the iPad by the vendor to meet patient and provider needs was also described by health care providers, as one health care provider noted:

> *That's all evolved, they've either added more—the nurses were finding what wasn't working and what the patients were finding that was more problematic that they took away and added and then they do the education of what has changed and how they need to change it.* [Health care provider]

The vendor also described how important it was to embed eQ Connect into health care providers' daily workflows and patients' daily routines; this was illustrated in this quote from a vendor:

> *The nurses have been very receptive because it is new and it's a change; there are things that we have to work through with the nurses in terms of fitting it in into how they do things and their workflow, and also making sure everyone's comfortable with the software.* [Vendor]

Being Influenced by Patient and Caregiver Characteristics

This theme emerged from the following subthemes: patients describing themselves as currently being stable and managing health issues, having tablet literacy or experiencing initial anxiety with eQ Connect, and being supported by a caregiver. Most patients described themselves as being stable at the moment and their condition as "hasn't gotten worse, hasn't gotten better," which was reiterated by their health care providers. Some of the patients had shared that they have been dealing with CKD and its manifestations, alongside other conditions (eg, high blood pressure and diabetes), for years and attributed not improving to the progression of their CKD. Patients had varying levels of tablet literacy, with some having previous experience and established comfort, as described by one patient as "I wasn't scared of using the iPad" and by another patient as "not a learning curve for me" and "I found it pretty simple to use"; others shared they had initial anxiety, but that with time and use they were able to use the tablet effectively.

This later finding was observed by the vendor who stated the following:

> *We find that some of our older patients have a little bit more time to get used to the technology, so for some of them it's a matter of just using it repeatedly until they get comfortable.* [Vendor]

Health care providers also noted the comfort level of patients using the tablet effectively, as described in the following quote:

> *It depends on the patient's aptitude with technology, on the demographic of the patient, and the patient population, because I think it is very easy to use but if the patient is a much older patient and obviously doesn't speak English, then it's kind of like one more thing the patient thinks, "Oh no, I don't want to have to bother with this.*" [Health care provider]

Supportive caregivers were also identified as a factor in using the tablet and, in some cases, the caregiver was entering the data on the tablet, as stated by a patient in the following quote:

> *My wife is doing that iPad because I cannot see the numbers, she inputs and I never touch that iPad.* [Patient]

Recommendations for Scalability

Codesign Locally, Embed Into Daily Routine and Workflow, and Deploy Over Time

Given how patients, caregivers, and providers valued the adaptability features of eQ Connect, future efforts to scale would benefit from a co-design approach. In our study, technological challenges (eg, entering data, small font and tablet size, and slow operating system) were resolved when brought to the attention of the vendor, who listened to feedback from the patients and health care providers. As one health system decision maker noted, future efforts "...need to think [about] the design approach, particularly how you set it up from the beginning to work the way it should for clinicians, data analysts, and patients using the app." It is also imperative to integrate eQ Connect into existing workflows and care processes for health care providers and the daily life of patients and their caregivers, so that it is comfortable, convenient, and efficient to use. This is illustrated in a quote by a health care provider who shares, "It just needs to be woven into the fabric of their daily routine, because otherwise, usually there's pushback when there's just

one more thing piled on top of their already hectic schedules." Further, a phased-in deployment over time, particularly to support the smaller size of the vendor organization, to ensure that local adaptation and co-design can occur is also recommended, as noted by a vendor who stated, "You wouldn't try to roll out to all different dialysis clinics across [the province]—it would be a continuing process, definitely be a gradual rollout adding patients and adding sites on to it."

Share the Benefits and Build the Case

Another key recommendation for the scalability efforts of eQ Connect is to share the benefits of the pilot and build the case for ongoing investment for the virtual care solution. This includes communicating the key functionalities of the virtual care solution, such as real-time monitoring, surveillance, and communication; reminders for inventory management; and accessible technological support and clinical expertise by *word of mouth* (ie, informally) and through *networks* (eg, Ontario Renal Network and media network), highlighting the benefits and experiences shared by patients. As one health care provider stated, "It would be awesome to get patients' input—there's a lot of positive feedback that I've heard from patients. Sharing the benefits, experiences of patients and nurses." Vendors and health system decision makers also described the importance of the need to strategically align the introduction of virtual care solutions, in this case eQ Connect, to broader health policy issues (ie, the economic, social, and health burden associated with CKD) and to key stakeholders' priorities (eg, Local Health Integrated Network [LHIN] and Patients First, Ministry of Health and Long-Term Care platform). Part of building the case will be to leverage the passion and expertise of leaders (ie, having super users is suggested). Support from leadership at all levels (eg, health care team, health care organization, and government) will be required, as will ongoing rapid-cycle evaluations to inform how best to sustain gains met and spread to other patients and their caregivers. As one health system decision maker noted, "We are really looking to scale now, and if it's not aligned with a LHIN priority or ministry priority, no one is going to care—so it's wonderful work, but it has to be part of a broader strategy."

Discussion

Principal Findings

This study delineated influencing factors and recommendations for scalability of a virtual care solution aimed at enhancing outcomes and experiences of CKD patients receiving PD at home. To our knowledge, this is one of the first empirical papers to elucidate factors that influence uptake and recommendations for scalability from multiple stakeholder groups.

Our study finding regarding how the hands-on training by the vendor is highly valuable is consistent with other literature that identifies using a hands-on learning approach as a key strategy for future usability of new technology [11]. In our study, the orientation provided to patients and caregivers was short in nature with minimal training that may reflect a well-designed product [36]. Similar to other studies using virtual care solutions [9,37,38], the ongoing accessibility and support from the CKD clinic team and vendor that patients had was key to using eQ Connect. Having virtual care solutions that are complimentary to, and integrated with, existing models of care is essential for usability and acceptance of new technology to support self-management and adherence [4,39]. In our study, the timely two-way exchange of data by patients, and in some cases their caregivers, with health care providers using eQ Connect made patients feel connected and supported. Trust between patients and health care providers and supportive communication between informal and professional caregivers are critical factors for the successful implementation of virtual care solutions for CKD and the provision of hemodialysis in the home environment [40,41].

Another prominent intervention mechanism, adapting the virtual care solution to meet the user's end needs and the recommendation to co-design locally in scalability efforts, adds to the existing literature base on the critical role of engaging and meeting the needs of end users [36,42-45]. User involvement in the design and development process is a fundamental human factors design principle and offers many benefits. Specifically, when users are involved, devices are created and adapted to reflect what users' needs are across their illness trajectories [36]. More recently, there are policy-level calls to include patients in identifying key virtual care strategies that will help the health system to be more patient-centric in nature [45]. Wider adoption of virtual care solutions requires processes to be redesigned from a patient-centric perspective to establish and sustain patient acceptance of care technology innovation [46]. In our study, there were variations around the health status of patients and tablet literacy that could have impacted the ongoing use and treatment adherence and were addressed over time by the CKD clinic team and vendor.

The co-design and adaptability of eQ Connect enabled technical and tablet literacy challenges to be addressed; they also enabled patients and their caregivers to continue to use the virtual care solution in their daily care routine and enabled an integration into the CKD clinic team's daily workflow. This finding exemplifies another human factor design principle around having an intuitive design that is embedded into existing care practices of patients and workflow of health care professionals [11,36,45,46] that should be less complicated and time demanding [9]. Other good design elements that emerged in a recent qualitative study include aesthetic appearance, practicality and ease of use, and supportive platforms to enable flow of data [36].

In addition to co-designing locally using a patient-centric approach, our study further highlighted the need to deploy over time and ensure that the vendor has the capacity to scale up the virtual care solution. This has implications for scalability of user-centric designs that require tailoring to specific contextual situations. Reconciling the trade-offs around upfront investment in innovation and future benefit and cost-efficiencies will require sharing the benefits of, and building a case for, continued investment in eQ Connect as an effective, patient-centric virtual care solution for CKD patients receiving PD at home. Specifically, sharing the positive impact eQ Connect had on participants, particularly patients and their caregivers and health care professionals, is paramount, as are continued efforts to strategically align this work with broader health policy issues

and key stakeholder priorities. Further research is required to examine and explore the outcomes and experiences using large-scale studies in different contexts associated with this patient-centric virtual care solution for CKD patients receiving PD at home.

Limitations

Study findings need to be interpreted with the following limitations. Given the study was conducted at two hospitals in Southern Ontario, the transferability of the qualitative data to other types of health care organizations may be limited. Selection biases may also have existed, as participants volunteered to participate in the study. There was a small sample size with a limited number of patients enrolled at each site to recruit into the study. Finally, the interviews were conducted during a 3-month window after the initial training session. This may have resulted in varying levels of familiarity with the app at the time of the interview. It is important to note that we do not feel this would bias our results and that interviewing parties with varying experience is important to capture issues relating to early usage while mitigating the impact of the learning curve.

Conclusions

Our study involved multiple stakeholder groups to elucidate the factors that influence uptake and recommendations for scalability of a virtual care solution for CKD patients receiving PD at home. Study findings can inform key stakeholders in their future efforts to enhance the uptake of implementation and scalability of virtual care solutions for CKD patients receiving PD at home that may also have relevance for other chronic diseases. Specific recommendations include the following: provide hands-on training and ongoing, timely support from the care team; co-design locally using a patient-centric approach adapting to meet user needs; embed into patients' daily routines and health care professionals' workflows; and deploy over time. Future efforts to scale up eQ Connect will require sharing the positive impact of eQ Connect and continued efforts to strategically align this work with broader health policy issues and key stakeholder priorities. Further research is required to examine and explore the outcomes and experiences using large-scale studies in different contexts associated with this patient-centric virtual care solution for CKD patients receiving PD at home.

Acknowledgments

The authors would like to acknowledge the Ontario Telemedicine Network for their engagement and operational oversight. The authors would also like to thank Canada Health Infoway and the Ministry of Health and Long-Term Care for funding this study.

The publication costs for this article have been covered by Canada Health Infoway Inc, a non-profit corporation funded by the Government of Canada. The funders had no role in study design, data collection and analysis, or decision to publish the manuscript.

Conflicts of Interest

None declared.

References

1. Saran R, Robinson B, Abbott K, Agodoa LYC, Bhave N, Bragg-Gresham J, et al. US Renal Data System 2017 Annual Data Report: Epidemiology of kidney disease in the United States. Am J Kidney Dis 2018 Mar;71(3S1):A7. [doi: 10.1053/j.ajkd.2018.01.002] [Medline: 29477157]
2. Webster G, Wu J, Terner M, Ivis F, de Sa E, Hall N. Canadian Organ Replacement Register Annual Report: Treatment of End-Stage Organ Failure in Canada, 2004 to 2013. Ottawa, ON: Canadian Institute of Health Information (CIHI); 2015 Apr. URL: https://secure.cihi.ca/free_products/2015_CORR_AnnualReport_ENweb.pdf [accessed 2019-03-04] [WebCite Cache ID 76d6VzRjV]
3. Vanholder R, Annemans L, Brown E, Gansevoort R, Gout-Zwart JJ, Lameire N, European Kidney Health Alliance. Reducing the costs of chronic kidney disease while delivering quality health care: A call to action. Nat Rev Nephrol 2017 Dec;13(7):393-409. [doi: 10.1038/nrneph.2017.63] [Medline: 28555652]
4. He T, Liu X, Li Y, Wu Q, Liu M, Yuan H. Remote home management for chronic kidney disease: A systematic review. J Telemed Telecare 2017 Jan;23(1):3-13. [doi: 10.1177/1357633X15626855] [Medline: 27269795]
5. Ong SW, Jassal SV, Miller JA, Porter EC, Cafazzo JA, Seto E, et al. Integrating a smartphone-based self-management system into usual care of advanced CKD. Clin J Am Soc Nephrol 2016 Dec 06;11(6):1054-1062 [FREE Full text] [doi: 10.2215/CJN.10681015] [Medline: 27173169]
6. Bello AK, Hemmelgarn B, Lin M, Manns B, Klarenbach S, Thompson S, Alberta Kidney Disease Network. Impact of remote location on quality care delivery and relationships to adverse health outcomes in patients with diabetes and chronic kidney disease. Nephrol Dial Transplant 2012 Oct;27(10):3849-3855. [doi: 10.1093/ndt/gfs267] [Medline: 22759385]
7. Liddy C, Rowan MS, Afkham A, Maranger J, Keely E. Building access to specialist care through e-consultation. Open Med 2013;7(1):e1-e8 [FREE Full text] [Medline: 23687533]
8. Tonelli M, Molzahn A, Wiebe N, Davison SN, Gill JS, Hemmelgarn BR, Alberta Kidney Disease Network. Relocation of remote dwellers living with hemodialysis: A time trade-off survey. Nephrol Dial Transplant 2015 Oct;30(10):1767-1773. [doi: 10.1093/ndt/gfv112] [Medline: 26038350]

9. Thilly N, Chanliau J, Frimat L, Combe C, Merville P, Chauveau P, et al. Cost-effectiveness of home telemonitoring in chronic kidney disease patients at different stages by a pragmatic randomized controlled trial (eNephro): Rationale and study design. BMC Nephrol 2017 Dec 05;18(1):126 [FREE Full text] [doi: 10.1186/s12882-017-0529-2] [Medline: 28381266]
10. Rucker D, Hemmelgarn BR, Lin M, Manns BJ, Klarenbach SW, Ayyalasomayajula B, et al. Quality of care and mortality are worse in chronic kidney disease patients living in remote areas. Kidney Int 2011 Jan;79(2):210-217 [FREE Full text] [doi: 10.1038/ki.2010.376] [Medline: 20927036]
11. Bello A, Molzahn A, Girard L, Osman M, Okpechi I, Glassford J, et al. Patient and provider perspectives on the design and implementation of an electronic consultation system for kidney care delivery in Canada: A focus group study. BMJ Open 2017 Dec 02;7(3):e014784 [FREE Full text] [doi: 10.1136/bmjopen-2016-014784] [Medline: 28255097]
12. Jain AK, Blake P, Cordy P, Garg AX. Global trends in rates of peritoneal dialysis. J Am Soc Nephrol 2012 Mar;23(3):533-544 [FREE Full text] [doi: 10.1681/ASN.2011060607] [Medline: 22302194]
13. Rubin HR, Fink NE, Plantinga LC, Sadler JH, Kliger AS, Powe NR. Patient ratings of dialysis care with peritoneal dialysis vs hemodialysis. JAMA 2004 Feb 11;291(6):697-703. [doi: 10.1001/jama.291.6.697] [Medline: 14871912]
14. Li J, Wang H, Xie H, Mei G, Cai W, Ye J, et al. Effects of post-discharge nurse-led telephone supportive care for patients with chronic kidney disease undergoing peritoneal dialysis in China: A randomized controlled trial. Perit Dial Int 2014 May;34(3):278-288 [FREE Full text] [doi: 10.3747/pdi.2012.00268] [Medline: 24385331]
15. Griva K, Yu Z, Chan S, Krisnasamy T, Yamin RBA, Zakaria FB, et al. Age is not a contraindication to home-based dialysis: Quality-of-life outcomes favour older patients on peritoneal dialysis regimes relative to younger patients. J Adv Nurs 2014 Aug;70(8):1902-1914. [doi: 10.1111/jan.12355] [Medline: 24495288]
16. Brown E. Peritoneal dialysis: Older patients report better quality of life than younger. Evid Based Nurs 2015 Jul;18(3):93. [doi: 10.1136/eb-2014-101989] [Medline: 25605820]
17. Kaldoudi E, Passadakis P, Panagoutsos S, Vargemezis V. Home care telematics for peritoneal dialysis: Field analysis and design considerations. In: Proceedings of the 5th International Conference on Information & Communication Technologies in Health. 2007 Presented at: 5th International Conference on Information & Communication Technologies in Health; July 5-7, 2007; Samos, Greece URL: http://iris.med.duth.gr/kaldoudi/wp-content/uploads/2015/05/Kaldoudi_Conf_19_ICICHT_2007_PD.pdf
18. Chaudhary K, Sangha H, Khanna R. Peritoneal dialysis first: Rationale. Clin J Am Soc Nephrol 2011 Feb;6(2):447-456 [FREE Full text] [doi: 10.2215/CJN.07920910] [Medline: 21115629]
19. Murphy SW, Foley RN, Barrett BJ, Kent GM, Morgan J, Barré P, et al. Comparative mortality of hemodialysis and peritoneal dialysis in Canada. Kidney Int 2000 Apr;57(4):1720-1726 [FREE Full text] [doi: 10.1046/j.1523-1755.2000.00017.x] [Medline: 10760108]
20. Chui BK, Manns B, Pannu N, Dong J, Wiebe N, Jindal K, et al. Health care costs of peritoneal dialysis technique failure and dialysis modality switching. Am J Kidney Dis 2013 Jan;61(1):104-111. [doi: 10.1053/j.ajkd.2012.07.010] [Medline: 22901772]
21. Lunney M, Lee R, Tang K, Wiebe N, Bello AK, Thomas C, et al. Impact of telehealth interventions on processes and quality of care for patients with ESRD. Am J Kidney Dis 2018 Oct;72(4):592-600 [FREE Full text] [doi: 10.1053/j.ajkd.2018.02.353] [Medline: 29699884]
22. Krishna V, Managadi K, Smith M, Wallace E. Telehealth in the delivery of home dialysis care: Catching up with technology. Adv Chronic Kidney Dis 2017 Jan;24(1):12-16. [doi: 10.1053/j.ackd.2016.11.014] [Medline: 28224937]
23. Hayashi A, Yamaguchi S, Waki K, Fujiu K, Hanafusa N, Nishi T, et al. Testing the feasibility and usability of a novel smartphone-based self-management support system for dialysis patients: A pilot study. JMIR Res Protoc 2017 Apr 20;6(4):e63 [FREE Full text] [doi: 10.2196/resprot.7105] [Medline: 28428168]
24. Raphael M, Nadeau-Fredette A, Tennankore K, Chan C. A virtual ward for home hemodialysis patients: A pilot trial. Can J Kidney Health Dis 2015;2:37 [FREE Full text] [doi: 10.1186/s40697-015-0072-7] [Medline: 26527130]
25. Kiberd J, Khan U, Stockman C, Radhakrishnan A, Phillips M, Kiberd BA, et al. Effectiveness of a Web-based eHealth portal for delivery of care to home dialysis patients: A single-arm pilot study. Can J Kidney Health Dis 2018;5:2054358118794415 [FREE Full text] [doi: 10.1177/2054358118794415] [Medline: 30210802]
26. Gordon EJ, Fink JC, Fischer MJ. Telenephrology: A novel approach to improve coordinated and collaborative care for chronic kidney disease. Nephrol Dial Transplant 2013 Apr;28(4):972-981. [doi: 10.1093/ndt/gfs552] [Medline: 23243040]
27. Osman MA, Okel J, Okpechi IG, Jindal K, Bello AK. Potential applications of telenephrology to enhance global kidney care. BMJ Glob Health 2017;2(2):e000292 [FREE Full text] [doi: 10.1136/bmjgh-2017-000292] [Medline: 29225932]
28. Weiner S, Fink JC. Telemedicine to promote patient safety: Use of phone-based interactive voice-response system to reduce adverse safety events in pre-dialysis CKD. Adv Chronic Kidney Dis 2017 Jan;24(1):31-38 [FREE Full text] [doi: 10.1053/j.ackd.2016.12.004] [Medline: 28224940]
29. Rojas SV, Gagnon M. A systematic review of the key indicators for assessing telehomecare cost-effectiveness. Telemed J E Health 2008 Nov;14(9):896-904 [FREE Full text] [doi: 10.1089/tmj.2008.0009] [Medline: 19035798]
30. Lew SQ, Sikka N. Telehealth and peritoneal dialysis in the US: Outcomes in practice. J Kidney Care 2018 May 22;3(3) [FREE Full text] [doi: 10.12968/jokc.2018.3.3.156]

31. Jeffs L, Jain AK, Man RH, Onabajo N, Desveaux L, Shaw J, et al. Exploring the utility and scalability of a telehomecare intervention for patients with chronic kidney disease undergoing peritoneal dialysis: A study protocol. BMC Nephrol 2017 May 10;18(1):155 [FREE Full text] [doi: 10.1186/s12882-017-0557-y] [Medline: 28486991]
32. RE-AIM.org. URL: http://www.re-aim.org/ [accessed 2019-02-04] [WebCite Cache ID 75y011tSj]
33. Berwick DM, Nolan TW, Whittington J. The triple aim: Care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
34. Barker PM, Reid A, Schall MW. A framework for scaling up health interventions: Lessons from large-scale improvement initiatives in Africa. Implement Sci 2016 Jan 29;11:12 [FREE Full text] [doi: 10.1186/s13012-016-0374-x] [Medline: 26821910]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Campling NC, Pitts DG, Knight PV, Aspinall R. A qualitative analysis of the effectiveness of telehealthcare devices (i) Are they meeting the needs of end-users? BMC Health Serv Res 2017 Dec 04;17(1):455 [FREE Full text] [doi: 10.1186/s12913-017-2408-8] [Medline: 28676054]
37. Ishani A, Christopher J, Palmer D, Otterness S, Clothier B, Nugent S, Center for Innovative Kidney Care. Telehealth by an interprofessional team in patients with CKD: A randomized controlled trial. Am J Kidney Dis 2016 Jul;68(1):41-49. [doi: 10.1053/j.ajkd.2016.01.018] [Medline: 26947216]
38. Narva AS, Romancito G, Faber T, Steele ME, Kempner KM. Managing CKD by telemedicine: The Zuni Telenephrology Clinic. Adv Chronic Kidney Dis 2017 Jan;24(1):6-11 [FREE Full text] [doi: 10.1053/j.ackd.2016.11.019] [Medline: 28224944]
39. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: Results of a patient survey and observational study. J Med Internet Res 2017 Dec 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
40. Rajkomar A, Farrington K, Mayer A, Walker D, Blandford A. Patients' and carers' experiences of interacting with home haemodialysis technology: Implications for quality and safety. BMC Nephrol 2014 Dec 11;15:195 [FREE Full text] [doi: 10.1186/1471-2369-15-195] [Medline: 25495826]
41. Jayanti A, Wearden A, Morris J, Brenchley P, Abma I, Bayer S, et al. Barriers to successful implementation of care in home haemodialysis (BASIC-HHD): 1. Study design, methods and rationale. BMC Nephrol 2013 Sep 17;14:197 [FREE Full text] [doi: 10.1186/1471-2369-14-197] [Medline: 24044499]
42. Aberger EW, Migliozzi D, Follick MJ, Malick T, Ahern DK. Enhancing patient engagement and blood pressure management for renal transplant recipients via home electronic monitoring and Web-enabled collaborative care. Telemed J E Health 2014 Sep;20(9):850-854 [FREE Full text] [doi: 10.1089/tmj.2013.0317] [Medline: 25046403]
43. Money AG, Barnett J, Kuljis J, Craven MP, Martin JL, Young T. The role of the user within the medical device design and development process: Medical device manufacturers' perspectives. BMC Med Inform Decis Mak 2011 Feb 28;11:15 [FREE Full text] [doi: 10.1186/1472-6947-11-15] [Medline: 21356097]
44. Lehoux P, Saint-Arnaud J, Richard L. The use of technology at home: What patient manuals say and sell vs what patients face and fear. Sociol Health Illn 2004 Jul;26(5):617-644 [FREE Full text] [doi: 10.1111/j.0141-9889.2004.00408.x] [Medline: 15283780]
45. Shaw J, Jamieson T, Agarwal P, Griffin B, Wong I, Bhatia RS. Virtual care policy recommendations for patient-centred primary care: Findings of a consensus policy dialogue using a nominal group technique. J Telemed Telecare 2018 Oct;24(9):608-615. [doi: 10.1177/1357633X17730444] [Medline: 28945161]
46. Boston-Fleischhauer C. The explosion of virtual nursing care. J Nurs Adm 2017 Feb;47(2):85-87. [doi: 10.1097/NNA.0000000000000444] [Medline: 28106680]

Abbreviations

CKD: chronic kidney disease
ESRD: end-stage renal disease
LHIN: Local Health Integrated Network
PD: peritoneal dialysis
RCT: randomized controlled trial
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance

Edited by G Eysenbach; submitted 27.12.17; peer-reviewed by S Ong, W Martinez, N Coulson; comments to author 20.03.18; revised version received 29.11.18; accepted 06.01.19; published 16.04.19.

Please cite as:
Jeffs L, Jamieson T, Saragosa M, Mukerji G, Jain AK, Man R, Desveaux L, Shaw J, Agarwal P, Hensel JM, Maione M, Onabajo N, Nguyen M, Bhatia R
Uptake and Scalability of a Peritoneal Dialysis Virtual Care Solution: Qualitative Study
JMIR Hum Factors 2019;6(2):e9720
URL: https://humanfactors.jmir.org/2019/2/e9720/
doi:10.2196/humanfactors.9720
PMID:30990460

Original Paper

Creating an mHealth App for Colorectal Cancer Screening: User-Centered Design Approach

Lauren Griffin[1], PhD; Donghee Lee[1], MA; Alyssa Jaisle[1], PhD; Peter Carek[2], MD, MS; Thomas George[3], MD, FACP; Eric Laber[4], PhD; Benjamin Lok[5], PhD; François Modave[6], PhD; Electra Paskett[7], PhD; Janice Krieger[1], PhD

[1]STEM Translational Communication Center, University of Florida, Gainesville, FL, United States

[2]Department of Community Health and Family Medicine, University of Florida, Gainesville, FL, United States

[3]Division of Oncology, College of Medicine, University of Florida, Gainesville, FL, United States

[4]Statistics Department, North Carolina State University, Raleigh, NC, United States

[5]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, United States

[6]Center for Health Outcomes and Informatics Research, Loyola University Chicago, Chicago, IL, United States

[7]Division of Cancer Prevention and Control, Department of Internal Medicine, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Lauren Griffin, PhD
STEM Translational Communication Center
University of Florida
Weimer Hall
University of Florida
Gainesville, FL, 32611
United States
Phone: 1 2198053886
Email: lngriffin@ufl.edu

Abstract

Background: Patients are increasingly using mobile health (mHealth) apps to monitor their health and educate themselves about medical issues. Despite the increasing popularity of such apps, poor design and usability often lead to suboptimal continued use of these apps and subsequently to poor adherence to the behavior changes at which they are aimed. One solution to these design problems is for app developers to use user-centered design (UCD) principles to consider the context and needs of users during the development process.

Objective: This study aimed to present a case study on the design and development process for an mHealth app that uses virtual human technology (VHT) to encourage colorectal cancer (CRC) screening among patients aged 50 years and above.

Methods: We have first provided an overview of the project and discussed its utilization of VHT. We have then reviewed UCD principles and how they can be incorporated into the development of health apps. We have described how we used UCD processes during the app's development. We have then discussed the unique roles played by communication researchers, computer scientists, clinicians, and community participants in creating an mHealth app that is credible, usable, effective, and accessible to its target audience.

Results: The principles of UCD were woven throughout the project development, with researchers collecting feedback from patients and providers at all stages and using that feedback to improve the credibility, usability, effectiveness, and accessibility of the mHealth app. The app was designed in an iterative process, which encouraged feedback and improvement of the app and allowed teams from different fields to revisit topics and troubleshoot problems.

Conclusions: Implementing a UCD process contributed to the development of an app, which not only reflected cross-disciplinary expertise but also the needs, wants, and concerns of patients.

(JMIR Hum Factors 2019;6(2):e12700) doi:10.2196/12700

KEYWORDS

communication; cell phone; mobile phone; culturally appropriate technology; interdisciplinary research; colon cancer; cancer screening

Introduction

Background

As technology improves, patients are increasingly using mobile apps to monitor their health and access medical information [1]. More mobile health (mHealth) apps are entering the market every year. However, poor development may diminish the usefulness of apps to patients [2]. Many mHealth apps are downloaded by patients but rarely used [3]. As such, it is recommended that teams developing mHealth apps use processes that consider the context and needs of users [4].

Over the past decade, access to the internet and smartphone ownership have increased to the point that virtually everyone in the United States has access to digital information. Furthermore, approximately three-quarters of Americans (including two-thirds of rural residents) have regular internet access [5]. Almost all Americans own a smartphone [6]. There is significant evidence that the penetration of the internet and mobile technologies could completely transform the way health care is delivered. It has the potential to effectively and efficiently deliver health behavior interventions with unsurpassed scalability [7-11]. Nonetheless, an expanding body of literature suggests that digital interventions lack the evidence-based standards required for apps to be usable in a health care setting or recommended for home use by health care providers [12-17]. The suggested reasons for lack of quality are lack of physicians' and patients' involvement in the development of such digital interventions. Although some recent research initiates strategies to involve stakeholders, this is not widespread yet [18].

This study presents a case study on the design and development process for an mHealth app that uses virtual human technology (VHT) to encourage colorectal cancer (CRC) screening among patients aged 50 years and above. Using participant observation, semistructured interviews, and document analysis, we have described the process by which a multidisciplinary team developed the app. The outcome was an mHealth app that reflects best practices across the medical, communication science, and computer science fields.

We have first provided an overview of the CRC screening project and discussed how it incorporates VHT. We have then reviewed the principles of user-centered design (UCD) and why UCD is useful for developing mHealth apps. We have described how the UCD process played out during the app's development, with a particular focus on how each set of researchers contributed to the overall design during each phase. In doing so, we expounded upon the unique roles played by communication scientists, computer scientists, clinicians, and community participants in creating an mHealth app that is credible, usable, effective, and accessible to its target audience. Our goal was to offer insights into the development process for other teams working on mHealth technology.

The Importance of Colorectal Cancer Screening

Among American men and women, CRC is the second leading cause of cancer death [19]. Racial and ethnic minorities are disproportionately impacted by CRC, with elevated incidences and mortality [20]. Although regular screening increases CRC detection and survival [21,22], minority patients face barriers such as time and monetary constraints and aversion to traditional screening procedures such as colonoscopies [23,24]. Similarly, rural patients are also disproportionately impacted by CRC morbidity and mortality [25]. Rural patients are less likely to understand the importance of screening and perceive cost as a barrier [26,27].

Fecal immunochemical testing (FIT) is a CRC screening procedure that may minimize the perceived barriers. Patients collect a stool sample at home and send it to a laboratory to test for microscopic blood that may indicate a tumor or colonic premalignant polyp. For patients at average risk, annual FIT is as effective as colonoscopy in detecting CRC [28,29]. As patients complete the test at home, FIT reduces barriers such as time, cost, and discomfort with colonoscopy. FIT is effective at increasing screening compliance for racial and ethnic minorities and rural populations [30].

Virtual Human Technology

VHT consists of computer-generated animated characters that can be used to communicate with people using speech or text [31]. VHT is increasingly common in health care. The technology has been used in studies on mental health care [32,33], assessing pain treatment [34-36], and patient and provider communication [37]. VHT has been used to increase patient satisfaction [37], improve the understanding of cancer risks [38], and give hospital discharge instructions [39]. The term *virtual human technology* is used specifically to describe three-dimensional human characters. This is different from an embodied conversational agent (ECA), which can be any anthropomorphic character, including a human. In other words, VHT is more specific than an ECA. VHT is also different from a chatbot, which is more general and includes all systems that can converse with users.

VHT may be useful for increasing CRC screening compliance for several reasons. Patients may feel more at ease discussing sensitive information because of VHT's sense of anonymity [40]. It may encourage patient disclosure [31], and it can also be used to provide tailored health information for patients, increasing perceptions of relevancy [41]. Similarly, demographic discordance between minority patients and providers is associated with worse medical outcomes [42-44]. VHT can match patients with demographically concordant virtual providers.

User-Centered Design

The design of an mHealth app impacts its use and effectiveness. As Schnall et al point out, many apps fail because they are not designed to meet the requirements of the people who are actually using them [4]. Such apps are unlikely to be used by patients [3]. Developing apps using a UCD process may address these shortcomings [3,4,45].

UCD is a multidisciplinary, iterative design process that involves actively seeking out and incorporating the feedback of users to ensure that tools are developed with a full understanding of their needs and requirements [46]. In UCD, social scientists act as translators between users and designers, using their research

skills to collect and interpret data about users and their needs [47].

The first phase in UCD is *needs investigation*. The goal of *needs investigation* is to identify potential users and learn about their specific needs for an mHealth app [3]. Many methods can be used in *needs investigation*, including cultural probes [48], interviews [49], and focus groups [3,4]. The second phase is *prototype development*. During *prototype development*, a trial version of the app is developed and tested, incorporating user feedback at multiple points [50]. The third phase is *evaluation*. During *evaluation*, researchers watch users test and evaluate the near-final app before rolling it out to larger audiences. Observing users can show researchers specifically how participants use the app and what problems they may experience [47]. These tests show researchers how the app functions when used by the type of people who will eventually use it on their own.

Although conceptually clear, in practice these phases are rarely clear-cut. As UCD is iterative, phases may blend together as researchers refine the app, troubleshoot problems, and seek additional feedback from users. This iterative process keeps the focus of development on users and ensures that the final product meets their needs [50].

Methods

First, we collected notes, meeting agendas, and other written documentation produced during the early stages of development. Second, the study's lead author engaged in participant observation of the development process, working as a postdoctoral researcher on the project while taking notes and working with the team on the app. Finally, the lead author interviewed 6 members of the development team about their role in the development process. The interviews were evaluative, approximately half an hour each, and transcribed for analysis.

A multiyear grant from the National Institutes of Health funded the development of the app. The design project is based at the University of Florida (UF), and the app will be a part of a clinical trial conducted at the UF Health Network, including Shands Hospital, launched in 2018. Furthermore, 3 core teams—clinical medicine, communication science, and computer science—contributed to the development of the app.

The app features an interaction with Agent Leveraging Empathy for eXams (ALEX) , a virtual human health care provider who educates patients about CRC screening and the benefits of FIT. During the clinical trial, we screened out patients who were at high risk of CRC (patients whose providers request more frequent colonoscopies or who have had colon cancer in the past) and those who were already within guidelines. Patients who are eligible for FIT see a series of tailored messages about CRC and its severity, their susceptibility to the disease, and how FIT can help them comply with screening guidelines. After visiting with ALEX, the app delivers an electronic message to patients giving them the option to request FIT from their primary care provider (PCP).

The app integrates into the UF Health Network and is delivered to patients directly through MyUFHealth (formerly known as MyChart), a Web-based medical portal. MyUFHealth lets patients securely access medical records, view laboratory results, and communicate with their PCP [51]. There are several advantages to integrating with MyUFHealth. First, using MyUFHealth to disseminate the app allows us to select patients with specific medical characteristics (ie, outside guidelines and average risk) for participation in the trial. Second, integrating with MyUFHealth lets us customize ALEX based on the demographic information in the patient's file. Finally, using MyUFHealth allows patients to quickly and securely request FIT from their PCP.

Results

Overview

The next section discusses how the UCD phases (needs investigation, prototype development, and evaluation) played out in the development of the CRC screening app. It focuses on the contributions of the communication science, computer science, and clinical teams to the credibility of the app, its usability, effectiveness, and accessibility. As UCD is iterative, many development processes happened simultaneously. The team often circled back to questions and concerns raised earlier in the process. Similarly, we sought and incorporated feedback from participants at multiple points in the development. As such, this section should be seen as a streamlined overview of the development process, which by necessity simplifies some elements.

Development Structure

We structured the development process around regular meetings between the 3 teams. The communication science team held weekly core meetings to coordinate development progress and integration into the larger university health system. The communication science and computer science teams met twice monthly to work on the hardware and software design of the app, with the communication science team providing feedback from potential users. The communication science and computer science teams also met with information technology (IT) representatives from UF Health as needed. We held these meetings in-person or online using a virtual meeting service. All 3 teams—communication science, computer science, and clinical—attended blended virtual and in-person meetings monthly and in-person meetings biannually.

This structure ensured that all teams understood how the app and clinical trial were evolving, even if they were not directly involved in a given branch of the work. It created flexibility for individual teams to meet as frequently as needed to accomplish their goals. Thus, individual teams could troubleshoot problems in a small-group setting and larger issues could receive input from all teams. We gained valuable feedback representing different disciplinary perspectives.

Phase 1: Investigating Needs

As the project began, teams addressed 3 foundational app components (1) the content of the app, (2) the integration between UF Health and the app, and (3) the app's software and user interface. During this phase, we developed the app

conceptually, tested acceptability to our target audience, and began creating the software.

Communication Science and Clinical Teams

The communication science team and clinical team began by identifying the medical content necessary for the app, specifically what it would need to convey to patients. The clinical team identified, through their experience with patients, common barriers to screening, including cost, time, and feelings of embarrassment caused by collecting a fecal sample. They paid specific attention to barriers that were common among minority and rural patients. To understand how clinicians address these barriers, the communication science team video-recorded a simulated conversation about CRC screening between a patient and clinician. A member of the clinical team played the role of the clinician and a member of the communication science team played the patient. The clinician described in lay terms the risks of CRC, the benefits of screening, and the biological changes that occur in older people, which raise the risk of CRC. This conversation formed the medical basis of script between the virtual human health care provider and the patient.

We also discussed the needs of clinicians and health care staff through over 50 interactions with the medical staff, including family medicine physicians, colorectal surgeons, health care administrators, patient navigators, and other players in the biomedical field. We asked questions about their processes and workflow when interacting with patients, incentives at the provider and practice levels for screening patients, and structural challenges in getting patients screened.

Through these interviews, we learned that physicians would likely welcome a tool to help them communicate about CRC with their patients. PCPs often have multiple topics to discuss with patients and limited time in which to do so. Providing patients with information about CRC before their appointment provides shared background for a conversation. Similarly, the amount of new information patients receive during an appointment can be overwhelming and stressful for patients, particularly those with lower health literacy. Providing some information beforehand reduces the amount of new information patients must absorb.

However, routine and regulation tend to govern medical environments. This means that physicians are unlikely to accept mHealth apps unless they fit into the regular workflow. mHealth interventions also cannot create extra work or take time away from patient care. These considerations informed the app's development. They are particularly important for the long-term dissemination of the app, as physicians and medical practices are a key channel for widespread distribution and adoption of the app by patients.

Computer Science Team

The computer science team began development of the virtual human health care provider. ALEX was created using Adobe Fuse, a design program, and Virtual People Factory, an interpersonal simulation system [52]. The computer science team created different versions of ALEX for focus group testing, designing a total of 8 characters varying along 3 dimensions: age (younger vs older), race (black vs white), and gender (man vs woman). They also had versions of the character in different attires, namely scrubs or business-casual office wear.

The computer science team began discussion of the hardware and software requirements of the app. With the larger team, they started the process of narrowing down which devices, browsers, and operating systems the app would support. As the app's target audience is older adults (aged 50 years and older), they also brought up questions of accessibility. This included the need for subtitles and clear audio to accommodate visual and hearing impairments. Similarly, the app interface needed to be understandable for people with limited smartphone experience. These conversations continued throughout the development.

Community Involvement

The communication science team conducted 8 focus groups (n=36) with potential users from January to May 2017. Participants were aged older than 50 years, and the team held groups broken down by race and gender with black men, white men, black women, and white women. They recorded, transcribed, and analyzed the focus group data qualitatively. This first round of focus groups provided the team with valuable information about the preferences, needs, and opinions of potential users before prototype development.

Discussion centered around 4 areas: health information seeking (*What features make health information trustworthy?*), initial thoughts on the virtual human (*Would you be comfortable talking to a virtual human about your health?*), CRC knowledge (*What words or feelings come to mind when you think about CRCs?*), and attitudes toward FIT (*What are your initial reactions to the FIT kit?*). During the discussion, moderators showed participants still photos of different versions of the virtual human health care provider. The most important finding was that participants were open to discussing their health with a virtual human health care provider, providing an essential rationale for proceeding with the app development.

Overall, Phase 1 provided information on patient and clinician user requirements for the app. It established, through community involvement, the general acceptability of using a virtual human health care provider to encourage CRC screening. It also generated insights into the technical requirements of the app and potential accessibility challenges.

Phase 2: Prototype Development

Computer Science Team

The computer science team had 2 main tasks during Phase 2: launching a working prototype of the app for user testing and planning the app's integration with MyUFHealth. Developing the prototype required multiple steps including the animation of the virtual human health care provider, coding the internal logic of the app (including options for randomization for the clinical trial), and designing the user interface. The computer science team and the communication science team met biweekly to discuss progress and address potential problems, creating an iterative workflow.

For instance, syncing voice actors' recordings of the script with the mouth movements of the virtual human health care provider required multiple iterations to reach an acceptable level. The communication science team originally asked colleagues in their college to serve as voice actors for a prototype ALEX. However, the varied speed and diction of nonprofessional voice recordings made it difficult for the computer science team to accurately sync the audio recordings with the lips of the virtual characters. To address this problem, the communication science team contracted professional voice actors to record the script. Paid voice actors recorded the scripts using professional equipment, which resulted in higher sound quality and greater syncing accuracy. The professional actors were also able to split audio files into segments to ease the process of syncing with the animation.

The computer science team began planning the app's integration with MyUFHealth. As MyUFHealth is an existing platform with its own constraints, the team was originally unsure whether it would be able to house the app entirely or whether it would be necessary to host portions of the intervention on an external server. Using an external site would allow for easier tracking of users but raised security concerns. Particularly problematic was the need to import demographic information—considered Protected Health Information (PHI)—into the app to customize the virtual human health care provider. Finally, it was decided that the app would be housed on its own secure server and users sent customized links with encrypted identification codes that allow us to track their movements and responses as they worked through the app.

Clinical Team

During Phase 2, the clinical team gathered information about programs ongoing in the UF Health Network to encourage CRC screening. They sought to understand what clinicians were currently doing to increase CRC screening so as to avoid designing an intervention that duplicates ongoing work. This is important both from a messaging perspective—ensuring that patients are not receiving competing messages—as well as from an experimental perspective. In evaluating the effectiveness of the app during the clinical trial, it is important to understand and avoid confounding influences to the greatest extent possible.

The clinical team also collected information about screening rates at the various clinic locations and within the different departments at UF Health. This information allows us to evaluate the effectiveness of the app by comparing past screening rates with screening rates during the clinical trial. It also helps us account for influences such as seasonal variation in screening rates.

Community Involvement

The communication science team conducted 13 focus groups (n=73) from November 2017 through August 2018. All participants were aged between 50 and 73 years. Owing to changes in the recruitment process, we separated some focus groups out by race and gender and others by gender only. Participants first filled out a questionnaire gauging their perceptions of CRC risk and screening. They then tested the prototype app on a Samsung Galaxy S7 smartphone provided to them by the moderators. After engaging with the app, participants filled out a second questionnaire examining their opinion of the app's technical aspects, the virtual human health care provider itself, and the CRC content. We recorded the focus groups and transcribed them for analysis.

The communication science team also held 38 think-aloud interviews during this timeframe, again using participants between the ages of 50 and 73 years. During think-aloud interviews, participants were asked to describe their thoughts and mental processes while using the app in real time [53]. The stream-of-consciousness data collected through think-aloud interviews let researchers see how participants are interacting with a tool, such as an mHealth app, in real time to better understand points of confusion and initial reactions.

Participants felt generally favorable toward the concept and script, with several indicating that they intended to ask their own PCP about FIT as a result of the experience. This provided preliminary evidence of the app's potential acceptability and effectiveness. However, participants were critical of the virtual human health care provider's appearance, indicating that the lack of a lab coat or medical name badge reduced the character's credibility. They also expressed concern about the look and movement of the virtual human health care provider. Many found the virtual human health care provider *creepy* and *unsettling*, with several saying that they averted their eyes from the character and listened to the voice instead of engaging visually.

In February 2018, we held a meeting of our Community Advisory Board, a group of patients, advocates, and professionals in the medical field. At the meeting, we sought feedback from the Community Advisory Board on the prototype version of the app and script. As with the focus groups, the Community Advisory Board members felt that the look and movement of the virtual human health care provider was unrealistic and distracting. They also gave feedback on the script's accessibility to those with lower literacy and/or health literacy and suggested areas within the script that needed to be expanded.

Communication Science Team

The communication science team incorporated the medical information collected during Phase 1 into a conversational script for the virtual human health care provider. They structured the conversation with ALEX around empirically-based constructs regarding CRC communication best practices. The original script identified 12 tailoring dimensions such as perceived susceptibility [54], perceived severity [55], perceived benefits [56], perceived barriers [23], self-efficacy [57], response efficacy [28], comparative risk feedback [58], risk probability [59], message source [60], narrative persuasion [61], demographic matching [62], and message framing [63]. Evidence suggests that these constructs can increase knowledge of cancer risks and screening and encourage behavioral change.

The team refined the script through input from multiple writers and readers, as well as the full app team and Community Advisory Board members. This led to significant changes, improving the script's flow and understandability. The team

also collapsed some constructs together for analytical purposes. Although the experimental design can accommodate multiple variables, analysis is complicated by each additional construct. The final message constructs are message source, susceptibility, severity, risk probability, response efficacy, benefits, barriers, narrative persuasion, and self-efficacy.

Phase 3: Evaluation

Communication Science Team and Community Involvement

In Phase 3, the communication science team adapted the script and messaging to reflect community preferences gleaned from Phase 2. They clarified the constructs within the script for ease of analysis in the clinical trial and sent the script to an expert at the American Cancer Society to read for clarify, accuracy, and comprehensiveness. These comments, as well as additional feedback from the clinical team, were used to finalize the script.

The communication science team also tested the near-final app with community members by conducting additional think-aloud interviews between September 2018 and January 2019. We held additional 7 focus groups and 15 think-aloud interviews. The total number of focus groups throughout the process was 28 (n=154), and the total number of think-aloud interviews was 53.

The think-aloud interviews initially revealed that significant problems remained with the appearance of the virtual human health care provider, particularly the black female version. To address these concerns, the computer science team created alternative versions of the black female character for testing by the communication science team with subsequent think-aloud and focus group participants. At this point, the development of the app became more intensively iterative, with the communication science team providing rapid feedback to the computer science team on changes that needed to be made to the app to achieve minimal acceptability from participants.

Computer Science Team

The computer science team refined the app during the evaluation phase, making changes as a result of community feedback, in particular, the results of the think-aloud interviews and focus groups. This involved discussions with the computer science team about potential changes in the graphic approach to the virtual human health care provider's appearance, moving from a more photorealistic look to one that was more stylized. The idea was that by going to a more stylized—but not cartoonish—look, participants would not be primed for photorealism and then put off by the limitations of the animation software and rendering process. Ultimately, the computer science team adapted models in Adobe Fuse to create a look that was somewhat stylized but also recognizable to viewers.

They also worked to integrate the app with MyUFHealth, ensuring that it was possible to demographically customize the virtual human health care provider for patients as per the study protocol. They paid particular attention to the need to track patients within MyUFHealth, as well as within the app itself, and the subsequent questionnaire (hosted on Qualtrics) and the need to link up these datasets for later analysis. They accomplished this through the aforementioned customized URLs and deidentification system. Using UCD principles helped ensure that the mHealth app we created was acceptable to patients along 4 major dimensions of user needs: credibility, usability, effectiveness, and accessibility.

Discussion

Principal Findings

By describing the creation of an mHealth app using UCD principles, we are able to better understand both the iterative nature of development when incorporating user feedback as well as the unique contributions of researchers across disciplines. Communication scientists, computer scientists, clinicians, and community participants all played specific and interrelated roles in ensuring that the final product was credible, usable, effective, and accessible for patients. We now summarize the specific components of these criteria and the contributions of each team in meeting them.

Credibility (Clinical, Communication Science, Computer Science, and Community Involvement)

Credibility had 3 main components: (1) accurate medical information, (2) association with the UF Health Network, and (3) a professional look and feel to the app design. Community members were ultimately the arbiters of what app features were and were not credible, as interpreted by the communication science team.

First, the communication science team worked with the clinical team during Phases 1 and 2 to create accurate content that reflects best clinical practices. This is in line with recommendations that health interventions be designed with input from subject matter experts [64]. Indeed, focus group participants in Phase 2 raised questions about the app's information source, with some explicitly asking whether UF Health was involved in development. Participants expressed skepticism about Web-based medical information, noting that such information is often misleading and inaccurate. However, they generally trusted the UF Health Network to provide them with credible information. Associating the app specifically with UF Health—a trusted medical provider—increased its credibility.

Second, the association between UF Health and trusted medical information was so strong that it carried over into participants' preferences for the look of the virtual human health care provider. The prototype app tested in Phase 2 had ALEX in a business-casual outfit, and there was no visible association with UF Health. Patients described this look as unprofessional and said that putting the virtual human health care provider in a lab coat would increase credibility. The computer science team made these changes for the think-aloud interviews and focus groups in Phase 3.

Third, participants said an app needed to have a professional look and feel to be seen as credible. Participants in Phase 2 focus groups and early Phase 3 think-aloud interviews expressed discomfort with the look and animation of the virtual human health care provider. A key theme was that participants wanted

the app to look like it was made by professional graphic designers to set it apart from other untrustworthy Web-based content. In other words, participants associated professional design and animation with medical credibility. Thus, even though clinical experts provided and vetted the app's content, it took the skills of the computer science team to make that expertise visible to participants.

Usability (Communication Science, Computer Science, and Community)

Usability had 2 main components: (1) intuitive app design and integration and (2) easily understood dialogue. As with credibility, community involvement helped operationalize these concepts in a way that reflected best practices from an academic perspective as well as from the perspective of the users themselves.

First, usability requires that the app design and interface be intuitive for both patients as well as clinicians and health workers. For patients, this meant that the app use and navigation needed to be self-explanatory even without instruction. Community feedback suggested a number of changes, which we incorporated into the app. For instance, the original working prototype had both a chat log and subtitles, which were seen as redundant. Similarly, although the app had a pause button, tapping the screen did not pause or play the interaction, which confused participants. Both these issues were corrected in the final version of the app.

For clinicians and health care workers, the app needed to intuitively fit into the clinical workflow to be usable, particularly with regard to requesting FIT. In designing this feature, the computer science team interfaced with UF Health to ensure that the appropriate medical professionals received the request through the appropriate channels, integrating with MyUFHealth. UF Health IT representatives indicated that clinical workers were accustomed to receiving information and requests from patients through the system. Using MyUFHealth, therefore, increased the usability of the app from the perspective of these employees.

Second, usability required that the app have understandable dialogue. This was a task taken up by the communication science team in translating the medical information from the clinical team into a coherent conversational script for ALEX. Multiple iterations of the script helped smooth out the sticking points in the dialogue, and feedback from a variety of readers increased cultural competency and eliminated jargon. Feedback from focus groups and think-aloud interviews suggests that these processes were largely successful—most participants felt that the app presented the information in an approachable and understandable way.

Effectiveness (Communication Science, Clinical, and Community Involvement)

Effectiveness had 2 main components: (1) increasing knowledge of CRC and screening and (2) changing behaviors. Preliminary results from focus groups suggest that the app meets these aims.

First, in designing the script for the virtual human health care provider, the communication science team sought feedback from the clinical team and community to establish what participants were likely to know about CRC and screening. This hands-on input supplemented the information in the health communication literature on knowledge of CRC. It helped strike a balance between providing too much information (overwhelming or boring patients) and providing too little (leaving patients with more questions than answers). For instance, some participants in the Phase 1 focus group did not know what CRC was, incorrectly conflating it with prostate cancer and assuming that only people with prostates need to be screened. To remedy this shortcoming, the communication science team revised the script to describe CRC as *colon cancer* or *cancer of the intestine*.

Preliminary feedback from the focus groups indicates that the app is effective at increasing knowledge of FIT testing and its appropriateness for CRC screening. Many participants did not know about FIT testing before the discussion and were unaware that there were alternatives to colonoscopy. Indeed, many expressed surprise that there was such an easy option available for screening. Other participants were unaware of the specific risks of CRC before engaging with the app.

Second, the communication science team drew on information from the health communication literature and the clinical team's expertise to write a script likely to change screening behaviors. For instance, both the literature and the clinical team stressed addressing barriers to screening, such as embarrassment about collecting a stool sample. To help lower these barriers and produce behavioral change, ALEX assures patients that they can complete the test in the privacy of their own home. This is important because messages that increase a person's self-efficacy—or how much they believe they can influence an outcome—are effective at changing behaviors. People are more likely to take action if they believe it is effective in reducing a threat.

Although we will not have quantitative data about the app's ability to produce behavioral changes until the end of the clinical trial, evidence from the focus groups suggests an increased desire to screen using FIT. Several participants asked how they could get FIT. Others explicitly stated a desire to use FIT, now that they knew it was effective. This suggests that the app will be effective at changing CRC screening behaviors.

Accessibility (Computer Science and Community)

Creating an app that is accessible to the target audience relied on 3 main considerations: (1) using the correct technology to reach the audience, (2) ensuring that the app is easy to find, and (3) making the app accessible to audiences with different abilities.

First, the computer science team balanced the need to reach a wide audience with the developmental challenges of creating an app supported by different devices, operating systems, and browsers. Community participants in the Phase 2 focus groups illustrated this need. Participants typically accessed MyUFHealth from their desktop computers rather than their mobile phones. Many participants use MyUFHealth infrequently, increasing the likelihood of forgetting their username and password. Resetting the password on mobile devices is clunky, so

participants defaulted to checking MyUFHealth from their desktop or laptop computers. Although we originally conceptualized the app as running mainly on mobile phones, the computer science team created a desktop version that increased the overall availability of the app for the target audience.

Second, the computer science team improved accessibility by integrating the app into the UF Health Network and MyUFHealth. Focus group participants expressed concern that they would be unable to find the app once we released it. By integrating the intervention into MyUFHealth, patients are able to log in to a system with which they are already familiar to access the app instead of downloading it from an unfamiliar Web-based source. Giving participants fewer tasks to complete before engaging with ALEX improves accessibility. Using MyUFHealth also allowed patients to view the intervention in the context of their relationship with their PCP and made requesting FIT easier as it could be done directly through MyUFHealth.

In addition, the app needed to be accessible to people who are hard of hearing and people with visual impairments. These requirements came out of the focus groups in Phase 1 and resulted in changes to the app's interface. The computer science team prioritized easily-read subtitles so that participants could easily follow along with ALEX, and we selected the voices for ALEX in part based on focus group feedback as to which were the clearest and most easily understood.

Conclusions

Ensuring that mHealth apps meet the needs of their target audience is an essential step toward widespread adoption. It is also a common shortcoming, with many mHealth apps being discarded by users shortly after initial usage owing to design failures that preclude their usefulness. Incorporating UCD principles into the design process of mHealth apps is one way to avoid this problem.

Our project used UCD principles in conjunction with expertise from communication science, computer science, clinical practitioners, and community members in an iterative process to create an mHealth app aimed at increasing CRC screening among adults aged 50 years and older. Through the phases of *needs investigation*, *prototype development*, and *evaluation*, we deliberately sought to highlight the opinions and concerns of community members as a way to increase the credibility, usability, effectiveness, and accessibility of the app. The overall product is one which aims to meet the needs of a variety of stakeholders as it moves through the clinical trial phase and into implementation across the health care system.

This study is not without limitations. A major limitation is lack of generalizability, with this project confined to 1 case study from the University of Florida. The iterative nature of UCD effected simultaneous collaboration among diverse academic disciplines, thereby presenting a potential challenge for replication in future research efforts where the culture and organizational structure may differ. However, stakeholder participation could be partially accomplished through centralized or remote participation, thus increasing the ability of other organizations that lack direct access to all key members to follow this blueprint.

Similarly, the study's design by necessity incorporated the perspectives of the participants and researchers themselves. Although we made all efforts to remain reflexive, it is possible that an outside observer would have drawn different conclusions, presenting a possible threat to validity. In particular, the iterative nature of UCD means that assumptions are continually challenged and revised throughout the development process. This means the perspectives of team members evolved throughout the project as more information was uncovered and incorporated. This paper captures the end point of these evolutions, but it also means that the process may have looked different depending on when the participants were interviewed. We do not believe this represents a significant threat to the overall utility of the paper in describing the UCD process but individuals wishing to incorporate similar processes in their own work should be aware of and open to similar changes in their own understandings.

Similarly, the utility of mHealth apps is largely dependent on the surrounding medical environments and patient characteristics, which may vary by institution and population. From a structural perspective, for instance, involvement of health care providers might be necessary to provide trainings for patients with low technical literacy to ensure successful application of the app in the real medical settings, requiring additional staff and resources. From a patient characteristics perspective, characteristics such as age, health status, health literacy, and technological literacy may impact uptake of mHealth interventions. Although these characteristics are important for widespread dissemination and utilization of mHealth technology, they are beyond the scope of this study to explore. Regardless, the benefits of using mHealth to foster lifesaving preventative care outweigh such potential challenges, particularly when interventions incorporate UCD principles.

Conflicts of Interest

None declared.

References

1. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]

2. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
3. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012 Sep;Suppl:49-56 [FREE Full text] [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
4. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
5. Pew Research Center. Demographics of Internet and Home Broadband Usage in the United States [Internet] URL: https://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2019-04-24] [WebCite Cache ID 77rqjQQOx]
6. Pew Research Center. Demographics of Mobile Device Ownership and Adoption in the United States [Internet] URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-04-24] [WebCite Cache ID 77rqqBFpr]
7. Azar K, Lesser L, Laing B, Stephens J, Aurora M, Burke L, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
8. Gold J, Pedrana A, Stoove M, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012 Feb 28;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]
9. Handel M. mHealth (mobile health)-using apps for health and wellness. Explore (NY) 2011;7(4):256-261. [doi: 10.1016/j.explore.2011.04.011] [Medline: 21724160]
10. Hswen Y, Viswanath K. Beyond the hype: mobile technologies and opportunities to address health disparities. J Mob Technol Med 2015 Jan;4(1):39-40. [doi: 10.7309/jmtm.4.1.9]
11. Pander T, Pinilla S, Dimitriadis K, Fischer M. The use of Facebook in medical education--a literature review. GMS Z Med Ausbild 2014;31(3) [FREE Full text] [doi: 10.3205/zma000925] [Medline: 25228935]
12. Breton E, Fuemmeler B, Abroms L. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
13. Cowan L, Van Wagenen SA, Brown B, Hedin R, Seino-Stephan Y, Hall P, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]
14. Eng D, Lee J. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
15. Hogan N, Kerin M. Smart phone apps: smart patients, steer clear. Patient Educ Couns 2012 Nov;89(2):360-361. [doi: 10.1016/j.pec.2012.07.016] [Medline: 22897983]
16. Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]
17. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
18. Modave F, Bian J, Rosenberg E, Mendoza T, Liang Z, Bhosale R, et al. DiaFit: the development of a smart app for patients with type 2 diabetes and obesity. JMIR Diabetes 2016;1(2) [FREE Full text] [doi: 10.2196/diabetes.6662] [Medline: 29388609]
19. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016 Jan 07;66(1):7-30 [FREE Full text] [doi: 10.3322/caac.21332] [Medline: 26742998]
20. Florida Statewide Cancer Registry. Florida Cancer Data System URL: http://fcds.med.miami.edu/inc/welcome.shtml [accessed 2019-04-19] [WebCite Cache ID 77kvJPT7p]
21. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med 1993 May 13;328(19):1365-1371. [doi: 10.1056/NEJM199305133281901] [Medline: 8474513]
22. Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013 Sep 19;369(12):1106-1114. [doi: 10.1056/NEJMoa1300720] [Medline: 24047060]
23. James A, Daley C, Greiner K. Knowledge and attitudes about colon cancer screening among African Americans. Am J Health Behav 2011 Jul;35(4):393-401 [FREE Full text] [Medline: 22040586]
24. Wang J, Moehring J, Stuhr S, Krug M. Barriers to colorectal cancer screening in Hispanics in the United States: an integrative review. Appl Nurs Res 2013 Nov;26(4):218-224. [doi: 10.1016/j.apnr.2013.08.005] [Medline: 24238084]
25. Anderson A, Henry K, Samadder N, Merrill R, Kinney A. Rural vs urban residence affects risk-appropriate colorectal cancer screening. Clin Gastroenterol Hepatol 2013 May;11(5):526-533 [FREE Full text] [doi: 10.1016/j.cgh.2012.11.025] [Medline: 23220166]
26. Davis T, Rademaker A, Bailey S, Platt D, Esparza J, Wolf M, et al. Contrasts in rural and urban barriers to colorectal cancer screening. Am J Health Behav 2013 May;37(3):289-298 [FREE Full text] [doi: 10.5993/AJHB.37.3.1] [Medline: 23985175]

27. Hughes AG, Watanabe-Galloway S, Schnell P, Soliman AS. Rural-urban differences in colorectal cancer screening barriers in Nebraska. J Community Health 2015 Dec 25;40(6):1065-1074 [FREE Full text] [doi: 10.1007/s10900-015-0032-2] [Medline: 25910484]
28. Duncan A, Turnbull D, Wilson C, Osborne JM, Cole SR, Flight I, et al. Behavioural and demographic predictors of adherence to three consecutive faecal occult blood test screening opportunities: a population study. BMC Public Health 2014 Mar 07;14(1):238 [FREE Full text] [doi: 10.1186/1471-2458-14-238] [Medline: 24606951]
29. Kapidzic A, Grobbee EJ, Hol L, van Roon AH, van Vuuren AJ, Spijker W, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014 Jul 1;109(8):1257-1264. [doi: 10.1038/ajg.2014.168]
30. Levy B, Daly J, Xu Y, Ely J. Mailed fecal immunochemical tests plus educational materials to improve colon cancer screening rates in Iowa Research Network (IRENE) practices. J Am Board Fam Med 2012;25(1):73-82 [FREE Full text] [doi: 10.3122/jabfm.2012.01.110055] [Medline: 22218627]
31. Lucas GM, Gratch J, King A, Morency L. It's only a computer: virtual humans increase willingness to disclose. Comput Human Behav 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
32. Albright G, Bryan C, Adam C, McMillan J, Shockley K. Using virtual patient simulations to prepare primary health care professionals to conduct substance use and mental health screening and brief intervention. J Am Psychiatr Nurses Assoc 2018;24(3):247-259. [doi: 10.1177/1078390317719321] [Medline: 28754067]
33. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Dec 16;7:42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
34. LaFond CM, Van Hulle Vincent C, Corte C, Hershberger PE, Johnson A, Park CG, et al. PICU nurses' pain assessments and intervention choices for virtual human and written vignettes. J Pediatr Nurs 2015 Jul;30(4):580-590. [doi: 10.1016/j.pedn.2015.01.022]
35. Wandner L, George S, Lok B, Torres C, Chuah J, Robinson M. Pain assessment and treatment decisions for virtual human patients. Cyberpsychol Behav Soc Netw 2013 Dec;16(12):904-909 [FREE Full text] [doi: 10.1089/cyber.2012.0707] [Medline: 23971429]
36. Wandner L, Heft M, Lok B, Hirsh A, George S, Horgas A, et al. The impact of patients' gender, race, and age on health care professionals' pain management decisions: an online survey using virtual human technology. Int J Nurs Stud 2014 May;51(5):726-733 [FREE Full text] [doi: 10.1016/j.ijnurstu.2013.09.011] [Medline: 24128374]
37. Schmid Mast M, Hall JA, Roter DL. Disentangling physician sex and physician communication style: their effects on patient satisfaction in a virtual medical visit. Patient Educ Couns 2007 Sep;68(1):16-22. [doi: 10.1016/j.pec.2007.03.020]
38. Persky S, Kaphingst K, Allen V, Senay I. Effects of patient-provider race concordance and smoking status on lung cancer risk perception accuracy among African-Americans. Ann Behav Med 2013 Jun;45(3):308-317 [FREE Full text] [doi: 10.1007/s12160-013-9475-9] [Medline: 23389688]
39. Zhou S, Bickmore T, Paasche-Orlow M, Jack B. Agent-User Concordance and Satisfaction with a Virtual Hospital Discharge Nurse. In: Intelligent Virtual Agents.: Springer International Publishing; 2014 Presented at: International Conference on Intelligent Virtual Agents; August 27-29; Boston, MA p. 528-541 URL: http://link.springer.com/chapter/10.1007/978-3-319-09767-1_63
40. Meeker D, Cerully J, Johnson M, Iyer N, Kurz J, Scharf D. SimCoach Evaluation: A Virtual Human Intervention to Encourage Service-Member Help-Seeking for Posttraumatic Stress Disorder and Depression. Santa Monica, CA: RAND Corporation; 2015.
41. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
42. Jacobs E, Rolle I, Ferrans C, Whitaker E, Warnecke R. Understanding African Americans' views of the trustworthiness of physicians. J Gen Intern Med 2006 Jun;21(6):642-647 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00485.x] [Medline: 16808750]
43. Laveist T, Nuru-Jeter A. Is doctor-patient race concordance associated with greater satisfaction with care? J Health Soc Behav 2002 Sep;43(3):296-306. [Medline: 12467254]
44. Schoenthaler A, Allegrante JP, Chaplin W, Ogedegbe G. The effect of patient-provider communication on medication adherence in hypertensive black patients: does race concordance matter? Ann Behav Med 2012 Jun 20;43(3):372-382 [FREE Full text] [doi: 10.1007/s12160-011-9342-5] [Medline: 22270266]
45. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May 01;4(3):666-684 [FREE Full text] [doi: 10.1177/193229681000400323] [Medline: 20513335]
46. Vrendenburg K, Mao J, Smith P, Carey T. A survey of user-centered design practice. 2002 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2002; Minneapolis, MN p. 471-478.
47. Frascara J. From user-centered to participatory design approaches. In: Design And The Social Sciences: Making Connections (contemporary Trends Institute). London, UK: Taylor & Francis Books Limited; 2002.

48. Park T, Chira P, Miller K, Nugent L. Living Profiles: an example of user-centered design in developing a teen-oriented personal health record. Pers Ubiquit Comput 2014 Aug 19;19(1):69-77. [doi: 10.1007/s00779-014-0812-1]
49. Raval MV, Taylor N, Piper K, Thakore M, Hoff K, Owens S, et al. Pediatric patient and caregiver preferences in the development of a mobile health application for management of surgical colorectal conditions. J Med Syst 2017 May 24;41(7). [doi: 10.1007/s10916-017-0750-3]
50. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004.
51. MyChart Central. 2018. MyChartCentral - FAQ Internet URL: https://www.mychartcentral.com/FAQ.aspx [accessed 2019-04-19] [WebCite Cache ID 77kuZDrvs]
52. Rossen B, Lind S, Lok B. Human-Centered Distributed Conversational Modeling: Efficient Modeling of Robust Virtual Human Conversations. In: Lecture Notes in Computer Science. Berlin: Springer; 2009 Presented at: International Workshop on Intelligent Virtual Agents; September 14-16; Amsterdam, The Netherlands p. 471-481 URL: http://link.springer.com/chapter/10.1007/978-3-642-04380-2_52
53. Jääskeläinen R. Think-aloud protocol. In: Gambier Y, van Doorslaer L, editors. Handbook of Translation Studies. Amsterdam: John Benjamins Publishing Company; 2010.
54. Shokar N, Carlson C, Weller S. Factors associated with racial/ethnic differences in colorectal cancer screening. J Am Board Fam Med 2008;21(5):414-426 [FREE Full text] [doi: 10.3122/jabfm.2008.05.070266] [Medline: 18772296]
55. Boonyasiriwat W, Hung M, Hon SD, Tang P, Pappas LM, Burt RW, et al. Intention to undergo colonoscopy screening among relatives of colorectal cancer Cases: a theory-based model. Ann Behav Med 2013 Dec 5;47(3):280-291. [doi: 10.1007/s12160-013-9562-y]
56. Christy SM, Perkins SM, Tong Y, Krier C, Champion VL, Skinner CS, et al. Promoting colorectal cancer screening discussion: a randomized controlled trial. Am J Prev Med 2013 Apr;44(4):325-329 [FREE Full text] [doi: 10.1016/j.amepre.2012.11.032] [Medline: 23498096]
57. Jerant A, To P, Franks P. The effects of tailoring knowledge acquisition on colorectal cancer screening self-efficacy. J Health Commun 2015;20(6):697-709 [FREE Full text] [doi: 10.1080/10810730.2015.1018562] [Medline: 25928315]
58. Weinstein ND, Atwood K, Puleo E, Fletcher R, Colditz G, Emmons KM. Colon cancer: risk perceptions and risk communication. J Health Commun 2004;9(1):53-65. [doi: 10.1080/10810730490271647] [Medline: 14761833]
59. Royak-Schaler R, Blocker D, Yali A, Bynoe M, Briant K, Smith S. Breast and colorectal cancer risk communication approaches with low-income African-American and Hispanic women: implications for healthcare providers. J Natl Med Assoc 2004 May;96(5):598-608. [Medline: 15160974]
60. Reinschmidt KM, Hunter JB, Fernandez ML, Lacy-Martinez CR, Guernsey de Zapien J, Meister J. Understanding the success of promotoras in increasing chronic diseases screening. J Health Care Poor Underserved 2006;17(2):256-264. [doi: 10.1353/hpu.2006.0066]
61. Hester C, Born W, Yeh H, Young K, James A, Daley C, et al. Decisional stage distribution for colorectal cancer screening among diverse, low-income study participants. Health Educ Res 2015 Jun;30(3):400-411 [FREE Full text] [doi: 10.1093/her/cyv006] [Medline: 25721254]
62. Sly J, Jandorf L, Erwin DO. Who's missing? Predictors of attrition following participation in culturally targeted educational breast and cervical cancer outreach programs for Latinas. J Health Commun 2015 May 26;20(7):851-858. [doi: 10.1080/10810730.2015.1018596]
63. Wilson C, Flight I, Turnbull D, Gregory T, Cole S, Young G, et al. A randomised controlled trial of personalised decision support delivered via the internet for bowel cancer screening with a faecal occult blood test: the effects of tailoring of messages according to social cognitive variables on participation. BMC Med Inform Decis Mak 2015 Apr 09;15:25 [FREE Full text] [doi: 10.1186/s12911-015-0147-5] [Medline: 25886492]
64. Campbell M, Fitzpatrick R, Haines A, Kinmonth A, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780]

Abbreviations

ALEX: Agent Leveraging Empathy for eXams
CRC: colorectal cancer
ECA: embodied conversational agent
FIT: fecal immunochemical testing
IT: information technology
mHealth: mobile health
PCP: primary care provider
UCD: user-centered design
UF: University of Florida
VHT: virtual human technology

Edited by G Eysenbach; submitted 22.11.18; peer-reviewed by K Sedig, J Marcotte; comments to author 24.01.19; revised version received 05.03.19; accepted 05.04.19; published 08.05.19.

Please cite as:
Griffin L, Lee D, Jaisle A, Carek P, George T, Laber E, Lok B, Modave F, Paskett E, Krieger J
Creating an mHealth App for Colorectal Cancer Screening: User-Centered Design Approach
JMIR Hum Factors 2019;6(2):e12700
URL: http://humanfactors.jmir.org/2019/2/e12700/
doi:10.2196/12700
PMID:31066688

Original Paper

Prototyping the Automated Food Imaging and Nutrient Intake Tracking System: Modified Participatory Iterative Design Sprint

Kaylen J Pfisterer[1,2], BSc, MSc; Jennifer Boger[1,2], BSc, MASc, PhD; Alexander Wong[1,2], BASc, MASc, PhD

[1]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

[2]Schlegel-UW Research Institute for Aging, Waterloo, ON, Canada

Corresponding Author:
Kaylen J Pfisterer, BSc, MSc
Department of Systems Design Engineering
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567
Email: kpfisterer@uwaterloo.ca

Abstract

Background: A total of 45% of older adults living in long-term care (LTC) have some form of malnutrition. Several methods of tracking food and fluid intake exist, but they are limited in terms of their accuracy and ease of application. An easy-to-use, objective, accurate, and comprehensive food intake system designed with LTC in mind may provide additional insights regarding nutritional support systems and nutritional interventions.

Objective: The aim of this study was to conduct a multistage participatory iterative design sprint of a Goldilocks quality horizontal prototype for the Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system. Specific design objectives included the following: (1) identify practice-relevant problems and solutions through user-centered participatory design, (2) mitigate feasibility-related barriers to uptake, and (3) employ user-centered technology development.

Methods: A 6-stage iterative participatory design sprint was developed and executed. A total of 38 participants and advisors representing 15 distinct roles (eg, personal support worker, nurse, and dietitian) were engaged in the design sprint. Subjective workload (Raw Task Load Index), subjective usability scales, and a modified Ravden checklist were used to assess project advisors' perceptions of the AFINI-T system prototype compared with the current method of food and fluid intake charting.

Results: The top priorities for this system were identified as the following: ease of use, high accuracy, system reliability, ease of maintenance, and requirement of integrating with the current PointClickCare system. Data from project advisors informed design decisions leading to a Goldilocks quality horizontal prototype of the AFINI-T system. Compared with the current food and fluid intake charting system, AFINI-T was perceived to have the following: less time demands ($t_{10.8}$=4.89; P<.001), less effort ($t_{13.5}$=5.55; P<.001), and less frustration ($t_{13.0}$=3.80; P=.002). Usability ratings of the AFINI-T prototype were high, with a subjective usability score mean of 89.2 and the highest ratings on a modified Ravden usability checklist of "very satisfactory" for 7 out of 8 sections.

Conclusions: The AFINI-T concept system appears to have good practice relevance as a tool for an intelligent food and fluid intake tracking system in LTC. The AFINI-T concept system may provide improvement over the current system, and advisors are keen to try the AFINI-T system. This research gives tangible examples of how the sprint method can be adapted and applied to the development of novel needs-based application-driven technology.

(JMIR Hum Factors 2019;6(2):e13017) doi:10.2196/13017

KEYWORDS

participatory iterative design; usability assessment; perceived workload; feasibility assessment; application-driven research; systematic prototyping; nutritional support; long-term care

Introduction

Background

The link between poor nutritional status and disease is well established; malnutrition is associated with decreased quality of life, increased hospital stays, pressure ulcers, morbidity, and mortality [1-3]. Furthermore, malnutrition-related costs the health care system US $10 billion per year in each the United States and United Kingdom [4,5]. Older adults are at increased risk of nutritional deficiency because of physical and physiological changes (eg, reduced lean muscle, less efficient gastrointestinal tracts, and changes in sensory ability such as smell or taste), in addition to having a higher degree of comorbidity [6]. Older adults living in long-term care (LTC) are particularly vulnerable; in Canada, 97% of older adults require assistance with activities of daily living (including eating assistance), 90% of the population is living with memory impairment, 61% of the population is on 10 or more medications, and 49% of the population is living with depression [7]; these demographics are similar in the United States [8]. On the basis of a recent Canadian study, approximately 44% of the LTC population is malnourished [9], which is consistent with a systematic review of global research (37 studies, 17 countries; malnutrition prevalence: 19% to 42%) [10]. Best practice metrics for ongoing nutritional assessment include monitoring unintentional weight loss, usual low intake of food, or other quality indicators to prioritize referrals and monitor effectiveness of nutritional support systems [11]. However, although inadequate intake is manageable [12], present guidelines for a nutritional intervention stipulate a resident must consume less than 75% of a meal most of the time [13-15]. Half of these residents who would benefit from an intervention are missed [14,15] because of difficulties assessing and charting food intake. Thus, monitoring nutritional status in LTC is crucial but difficult to do so effectively.

In LTC, nursing assistants or personal support workers (PSWs) chart food and fluid intake of residents using either a paper-based or an electronic form to capture intake across a meal at 25% incremental proportions of intake. The accuracy of these methods is known to be poor, with incorrect estimates over 50% of the time [16]. One contributing factor is time constraints in the LTC environment, and it is further confounded by frequent retrospective charting, which increases the probability of reporting errors [13]. Although accuracy is important to ensure appropriate referrals of residents to a registered dietitian (RD) [14], the current method fails to differentiate among aspects of a meal; equal consumption across a plate is assumed. To address this, Andrews and Castellanos developed a food-type specific tool; however, consumption was still underestimated 25% of the time [13]. The challenge remains that comparisons either require time-consuming methods or need to be completed by highly qualified personnel [14].

Technological innovations may provide a solution to remove subjectivity, enhance reproducibility, and inform higher levels of detail. There has been some progress in automatic food intake tracking systems. For example, several devices have been proposed for an individual to track and manage weight loss by recording intake using a mobile device [17-20]. Although this on-the-go approach could potentially be modified for appropriate use in LTC settings, in its current state, it is tailored for a different purpose, relies on self-monitoring, and does not adhere to related best practices for food and fluid intake. In addition, they require a series of images from multiple perspectives [17] or depend on reference objects to infer scale (ie, fiducial marker) [19]. In a time-constrained environment such as LTC and hospital settings, these requirements make these approaches infeasible. Consistent with this apparent gap, a 2016 review by Pouladzadeh et al [20] summarizes both traditional and newer (smartphone vision-based) methods for calorie intake tracking in the context of weight loss and weight maintenance. They conclude that several challenges remain, including the following: the explicit need for user acceptance studies of nutritional monitoring technology, consideration of more complex meal scenarios, and computational requirement consideration [20]. Within the LTC context, the closest technological solution was a comparison to estimate food waste of regular- and modified-texture diets either with the visual estimation method or by using digital photographs afterward [21].

Objectives and Goals

The above highlights the need for an easy-to-use, accurate, and comprehensive food intake system designed with the LTC context in mind. The goal of this research was to collaborate with representative end users to design a novel prototype system for Automated Food Imaging and Nutrient Intake Tracking (AFINI-T). End users in this context were team members working in LTC, involved in monitoring resident food intake (eg, PSWs and RDs). We developed a Goldilocks quality horizontal prototype by accomplishing the following objectives: (Objective 1) identify practice-relevant problems through user-centered participatory design, (Objective 2) remove feasibility-related barriers to uptake, and (Objective 3) facilitate confidence in design decisions for user-centered technology development. Our guiding principle was to accelerate research to uptake of novel technological solutions through practice-informed research. Each of the 3 objectives outlined above had several goals as follows: (Goal A) understand workflow and the problem space including user perceptions of workload of the current system (Objective 1); (Goal B) conduct a needs assessment within the problem space (Objective 1); (Goal C) establish functional criteria for usability and feasibility, including user interface requirements (Objective 2); (Goal D) evaluate a user-driven, practice-relevant early-stage prototype to inform future directions, including user perceptions of workload, usability, and receptivity of the AFINI-T system prototype (Objective 3). The primary contribution of this study is the novel AFINI-T system design created through the participatory iterative design by (1) the identification of functionality requirements and design considerations, (2) the findings and insights from user testing, and (3) a demonstration of and reflection on the effectiveness of this participatory iterative design methodology with a multidisciplinary team of project advisors. The remainder of this paper is organized as follows: the combined Design Stages section presents the 6 stages used in the design process, along with related results and

discussion for each stage, followed by a general discussion before closing with overreaching conclusions.

Methods

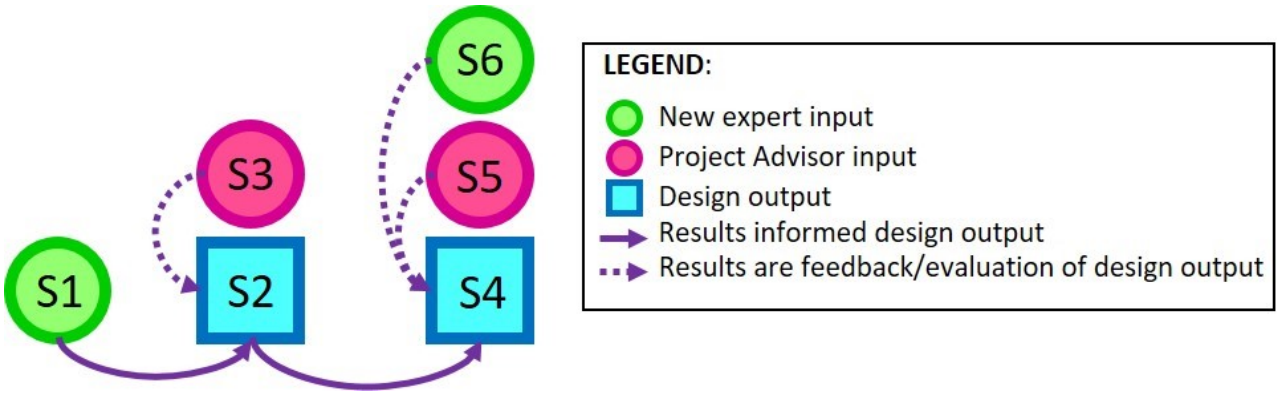
Overview

Our goal was to create a Goldilocks quality horizontal prototype. "Goldilocks quality" refers to having the "just right" amount of fidelity to elicit useful feedback from users without having to build an entirely functional prototype [22]. A horizontal prototype refers to a user interface–based design to allow user feedback on an early-stage conceptual walk-through of the process [23]. We implemented an iterative participatory iterative design process, modeled off the Google Sprint framework, to develop and evaluate this prototype for monitoring food and fluid intake in LTC [22,24]. The 6 stages of our process were the following:

1. *STAGE 1:* Design Ideation
2. *STAGE 2:* Reflect and Storyboard (*see Multimedia Appendix 1)*
3. *STAGE 3:* Storyboard Critiques (*see Multimedia Appendix 1)*
4. *STAGE 4:* Design of the Goldilocks Quality Horizontal Prototype
5. *STAGE 5:* Usability Assessment
6. *STAGE 6:* Final Validation

The design process was guided by several conceptual frameworks: (1) conducting interdisciplinary research [25,26], (2) leveraging user-centered design and participatory design [27,28], (3) applying rapid prototyping methodology via a modified Sprint [22,23]; (4) applying best practices for user interface design [23,29-33]; and (5) evaluating usability [34,35] and perceived workload [36]. The flow of information through each stage is shown in Figure 1. For brevity, the methods (including collaborators, data captured, and analyses), results, and discussion for Stages 1 and 4 to 6 are presented below, within the context of each stage; details regarding Stage 2 and 3 can be found in Multimedia Appendix 1 [37-43].

Figure 1. An overview of the 6 stages (eg, S1 means Stage 1), including information flow between stages. Solid arrows indicate results directly influencing design output (eg, S2's story boards and S4's Goldilocks prototype). Dashed arrows indicate feedback on a design stage. Feedback was collected from expert input (S1, S6 in green) and from ongoing project advisor engagement input (S3, S5 in pink).



STAGE 1: Design Ideation – Methods

The purpose of Stage 1 was to engage with end users as collaborators to establish design directions. Specifically, we sought to understand the current workflow, evaluate priorities, understand the perceived workload of the current system, and identify potential project advisors. The output from this directly informed Reflect and Storyboard (Stage 2) and Usability Assessment (Stage 5).

Participants

Stage 1 comprised a 60-min workshop in which 3 activities were completed: *Activity 1: The "Ask the Experts" activity*; *Activity 2: Priority ranking survey completion*; and *Activity 3: "Vote with dots" exercise* to keep participants engaged and reflect on priorities. A total of 3 research assistants, plus the lead author, took notes during this discussion and transcribed several comments verbatim. Following the workshop, 3 informal open-ended interviews were conducted to further inform the problem space. The lead author took notes during these interviews; several comments were transcribed verbatim.

For the workshop, 21 participants representing 12 LTC and retirement homes were recruited through self-enrollment with the following roles: Administrative Assistant, Chef, Dining Lead (similar to a dining room manager), Director of Recreation, Dietary Aides, Neighborhood Coordinator, Recreation Assistant, Restorative Care, Senior Nurse Consultant, Directors and Assistant Directors of Food Services, Nurse, and PSWs. Activities were discussed with the Schlegel-UW Research Institute for Aging's (RIA) Research Application Specialist for input on how to successfully conduct this exercise with front-line team members.

Tools

Activity 1: The "Ask the Experts" Activity

Workshop participants were asked about their experience with food and fluid intake. This aimed to build participants' confidence in the value of their experiences while probing current workflow and problem space.

Activity 2: Priority Ranking Survey

Participants independently completed a survey to evaluate priorities and needs to limit bias. This survey asked about the

current charting process (eg, when it is done, task completion time, and barriers and facilitators to task completion). For evaluating priorities, 5-point Likert scales were used to rate 16 statements' importance from "Not Important" (ie, 0) to "Very important" (ie, 4) or "Not Applicable." Perceived workload of the current system was retrospectively evaluated with the Raw Task Load Index (RTLX) [36,44] for its application simplicity and comparability to the NASA-TLX [44-47].

Activity 3: "Vote With Dots" Exercise

Modeled from the study by Knapp et al [22], participants transposed their individual Activity 2 responses into a group response by voting their preference using stickers on giant sticky notes to amalgamate opinions, keep participants engaged, and facilitate additional discussion.

Statistical Analysis

Given the nature and size of this pilot study, a preliminary thematic analysis was used for qualitative components (eg, discussions, comments, and verbal/written feedback) that was combined with descriptive statistics for quantitative information, including the average (μ), SD($_\sigma$), mode, and median scores [48]. For scales with 5 or more categories (eg, RTLX), μ($_\sigma$) is used; the mode was used for categorical data with fewer than 5 categories (eg, Ravden Checklist). A weighted average was used to analyze Likert survey questions, excluding "Not Applicable," to yield a ranking of each statement.

STAGE 4: Design of the Goldilocks Quality Horizontal Prototype – Methods

The purpose of Stage 4 was to create low-fidelity prototypes by incorporating the most promising solution concepts identified through the storyboard critiques in Stage 3. These prototypes were then used for pilot evaluation in Stage 5's usability assessment.

Tools

Design decisions were informed by heuristics, as in Stage 2 [23,32,33], and feedback received from the storyboard critiques in Stage 3. The following heuristics were emphasized: universal usability was considered by testing the prototypes with different types of users (eg, academics and PSWs), providing informative feedback and error prevention, the output in this stage (Stage 4) was a Goldilocks quality horizontal prototype. This included interfaces for each of the 3 levels of primary users currently involved in residents' food and fluid intake charting (ie, PSW, registered nursing team, and RD).

STAGE 5: Usability Assessment – Methods

The goal of Stage 5 was to elucidate preliminary feasibility early on with end users through the evaluation of prototypes through pilot testing. The output from this stage informed how the prototypes could be improved for the development of a working system in the future.

Prototypes were evaluated by comparing perceptions of the AFINI-T prototype with the system currently in place with regard to usability and workload. Usability was assessed using the Subjective Usability Scale (SUS) [34] from the user perspective and a modified Ravden usability evaluation checklist [35] from technical experts' perspectives; items pertaining to help, including all of section 9, were removed, as this was beyond the scope of the Goldilocks quality horizontal prototype.

Participants

A total of 4 project advisors from Stage 4 were tester participants (PSW, Dining Lead, Dietary Aide, and Nutrition Research Expert). By word of mouth, 2 new project advisors requested inclusion as observers for a total of 6 advisors. All testing was completed in person though one-on-one sessions. Testing sessions were audio-recorded and relevant quotes were transcribed verbatim. Testing began with an interview walk-through of the prototypes based on the script adapted from a study by Knapp et al [22] to ascertain usability and feasibility barriers. A novel, predefined strict set of tasks was completed by each advisor. The student investigator completed a checklist to capture the degree of success to which each task was completed (ie, success, required prompting, or failed).

Tools

The RTLX [36,44] was administered to enable comparison of perceived workload of the current method in place with the AFINI-T system prototype (Table 1). Usability was assessed with the SUS, which was selected over other usability questionnaires for its ease of use, minimal training requirements, and low application time [45,49]. The RTLX and SUS were also completed by the 2 observers (Director and Assistant Director of Food Services) based on their experience during the observation. These 2 project advisors had no previous experience or knowledge of this project.

For evaluating usability more formally, an adapted Ravden checklist was used by 2 technical experts with backgrounds in systems design engineering and limited exposure to the users' perspectives. The Ravden checklist was selected for its low cost and ease of use to assess the interface with good interrater reliability and predictive validity [45,49] (Multimedia Appendix 3).

Statistical Analysis

A 2-tailed *t* test assuming unequal variances [50,51] was conducted to compare the current system and the AFINI-T system for users' perceived workload for the RTLX. Quantitative data were analyzed using descriptive statistics, with highlights from qualitative data as described in Stage 1.

STAGE 6: Final Validation – Methods

The goal of Stage 6 was to receive additional feedback from a group of RDs, directors, and assistant directors of food services to provide a fresh perspective to minimize bias.

Participants

The RDs, directors, and assistant directors of food services from across the Schlegel Villages were invited to participate in a webinar outlining the progress to date, along with tandem survey completion for assessing perceived usability and workload. A total of 13 people participated in the webinar (43% participation rate), which is consistent with the typical attendance of quarterly dietitian meetings at Schlegel Villages because of scheduling complexities.

Results

STAGE 1: Design Ideation – Results

Results from Stage 1 pertained to Objective 1: address a practice-relevant problem through user-centered participatory design (Goals A and B) and Objective 2: remove feasibility-related barriers to uptake and are as follows (Goal C):

Goal A: Understand Workflow and Problem Space

PSWs, registered nursing team, and RDs are primary users who conduct charting of food and fluid intake on iPads. This charting is completed whenever primary users have time, which could be during meal service or retrospectively, consistent with the study by Andrews and Castellanos [13]. In a follow-up discussion with the organization-wide director of food services, who is responsible for policy, she indicated that conducting food intake in real time is mandated (as opposed to retrospectively), but from the workshop discussion, it is clear there is a gap between policy and practice. Although the workflow of AFINI-T is congruent with this mandate, a solution to support this mandate in practice may require policy modifications. For example, a person may need to be assigned to the sole task of tracking food and fluid intake during mealtime, which means he or she would be unavailable to provide assistance with residents' care needs for the duration of the meal. Changing policy is outside the scope of the current AFINI-T project but having sensitivity to this issue provides helpful context and informs that this may be a potential barrier to uptake of the system in practice.

Regarding the current system, respondents appreciated the ability to track fluids, so they need not manually add, and the output has units (mL). Although the current system is dependable, substantial barriers and limitations were identified regarding the effectiveness and accuracy of the current system. A workshop participant shared:

> *What's being collected for solid food isn't useful. It's so high level and minimal can't make use of it. [We] can't infer anything regarding health or category of at-risk. [We] look at last 7 days, see "they had 75% of a meal so they're eating well", but it doesn't say anything. [We] don't get a lot of info from the charts.*

Insufficient time, data inaccuracy, unreliability, and nonstandardized measurements were identified as the largest barriers for task completion. In addition, the inability to differentiate among types of foods and lack of relation to original serving size lead to data interpretation difficulties. For example, some residents prefer half portions; if they eat half of their portion, this could be recorded as 50% (ie, half of the serving they received) or it could be input as 25% (ie, one-fourth relative to the full portion). There is no guarantee that the proportion is input accurately or consistently. These themes were apparent through 2 sources, the "Ask the Experts" as well as on the survey. For more detail regarding the current system's retrospective analysis of perceived user workload, see the sections of Table 1 pertaining to the "Current" system.

Table 1. Comparing retrospective perceived users' workload measures of current food/fluid intake system from Stage 1 to the Automated Food Imaging and Nutrient Intake Tracking prototype results from Stage 5.

Workload demand[a] and system	Mean	Mode(s)	Minimum	Maximum	Responses, N	*t* test (df)	*P* value
Mental demand							
Current	10.2	6	4	19	10	2.56 (13.8)	.023
AFINI-T[b]	4.4	3	1	10	6	2.56 (13.8)	.023
Physical demand							
Current	6.4	2	1	15	9	1.41 (12.5)	.183
AFINI-T	3.5	1	1	6	6	1.41 (12.5)	.183
Time demand							
Current	16.7	19	5	20	10	4.89 (10.8)	<.001
AFINI-T	5.5	3	1	12	6	4.89 (10.8)	<.001
Performance							
Current	15.2	18, 20	3	20	10	0.722 (13.7)	.722
AFINI-T	16.8	20	11	20	6	0.722 (13.7)	.722
Effort							
Current	13.2	6	6	20	10	5.55 (13.5)	<.001
AFINI-T	3.7	3	1	7	6	5.55 (13.5)	<.001
Frustration							
Current	11.5	15	1	20	10	3.80 (13.0)	.002
AFINI-T	3	2	1	8	6	3.80 (13.0)	.002

[a]Values could take on a range from 0 to 20; 0 implies no workload and 20 implies highest imaginable workload except in the case of performance which is reverse coded.

[b]AFINI-T: Automated Food Imaging and Nutrient Intake Tracking.

Goal B: Conduct a Needs Assessment of Problem Space Including Priority Areas

Workshop participants were asked to rate need statements' importance. The top 3 ranked priorities were tied among (1) "ease of use" and "accuracy" (μ=3.9, mode: "very important," 15 out of 16 votes), (2) "reliability" and "maintenance" (μ=3.9, mode: "very important," 14 out of 16 votes), and (3) "The system should work well with PointClickCare" (μ=3.8, mode: "very important," 12 out of 16 votes).

The following 5 themes emerged as wishes for a novel system to extend beyond the current infrastructure: (1) being able to leverage weight of food as a ground truth instead of relying solely on subjective proportions, (2) having the ability to track trends over time, (3) being able to discriminate among types of food, (4) being able to include fluid intake as well to discriminate between types of fluids, and (5) operating the system in different modes to accommodate various use cases (ie, in the dining room vs for in-room service). One additional, complementary theme relevant to priorities, identified independently through 3 interviews, was the need to support prioritizing referrals that consider symptoms and risk flags' severity. One project advisor articulated:

> *There is 1 Registered Dietitian for 300 residents. It's impossible to track properly … People are often missed because nurses aren't identifying properly… If charting were accurate, this would help with the referral process.*

Goal C: Establish Functional Criteria for Usability and Feasibility

The current system mode time to complete the task defined the time completion target: 10 to 14 min, maximum, per neighborhood (ie, "ward") comprising 16 residents. Of the 21 workshop attendees, 11 self-identified as being involved in charting resident food and fluid intake and were asked about the amount of time required to complete intake charting for each type of food, fluid, or snack. Survey responses are outlined in Table 2.

Table 2. Summary of length of time required to complete food and fluid intake charting for 1 neighborhood comprising 16 residents (Stage 1).

Charting type	Mode time (min)	Responses[a], n/N (%)	Time range (min)
Food (meal)	10 to 14	3/9 (30)	<10 to 25+
Fluid	10 to 14	4/10 (40)	<10 to 25
Snack	<10	5/9 (64)	<10 to 19

[a]n is the number of responses with the mode rating out of N, the total number of responses.

STAGE 4: Design of the Goldilocks Quality Horizontal Prototype – Results

Design heuristics were applied in the 4 ways, and sample output from this stage is illustrated in the right pane of Figures 2 and 3 with additional inspiration from commercially available online health care tools (Multimedia Appendix 2) First, related to universal usability, mapping was considered through matching the system with users' language and familiar concepts in reality (eg, Figure 2 contains tab names for snacks, such as "AM," "PM," and "HS", which refer to the morning, afternoon, and evening snacks, respectively) [23,32]. Second, informative feedback on a change of state was provided [23,33] when users attempted to submit or track an action; there is a pop-up banner at the bottom of the screen (not shown). Third, error prevention [23,32,33] was incorporated through limiting types of responses and providing feedback. For example, the PSW interface would prompt for a picture or a progress note before submission, with the ability to finish charting at a later point of the meal service. Fourth, efforts were made to reduce short-term memory load and enhance visibility/discoverability [23,32,33] by placing the workspace into panes, with all information accessible on 1 screen. Other features included making "smart" suggestions when selecting items or filling out portion sizes. For example, notes entered from the RD interface (not shown) would auto populate on the RD instructions tab in the PSW interface.

Figure 2. Stage 2 personal support worker user interface. Output from Stage 3 included a heat map on the most promising aspects (red indicates more votes, n=5) with qualitative feedback highlights for additional considerations. The right pane illustrates an example of the prototype interface. Numbers correspond to the flow of information and adapted feedback from Stage 2 through to 3 and 4 using the first example (#1 in pink) to further illustrate flow with the dashed arrow.

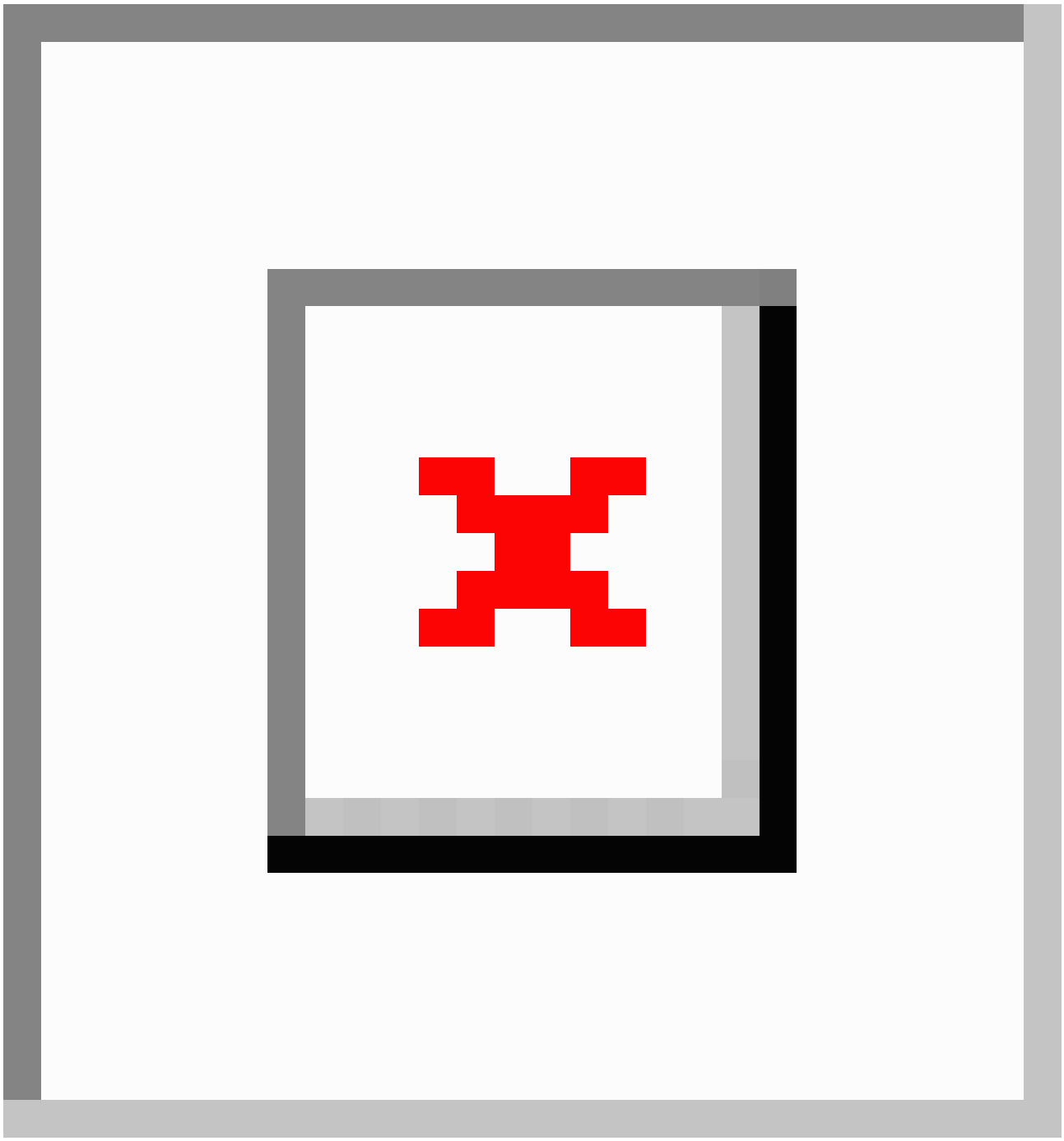
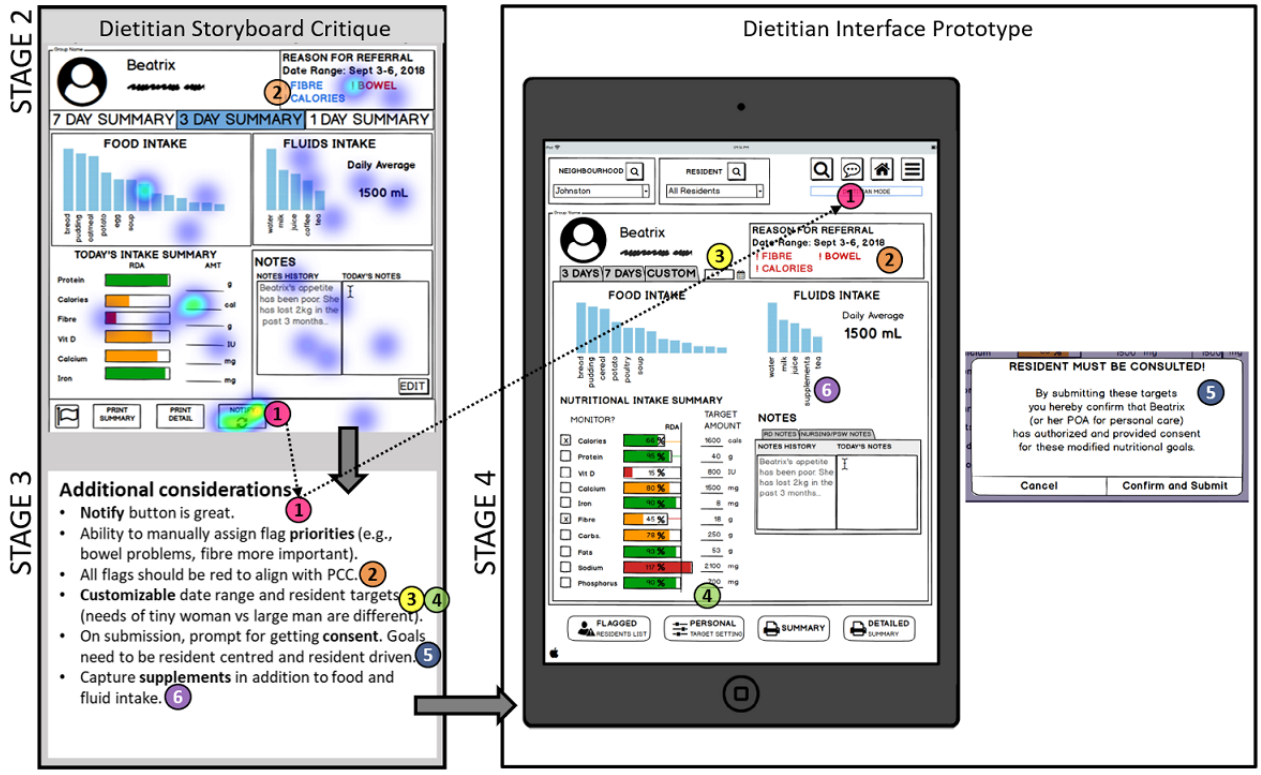


Figure 3. Stage 2 Registered Dietitian user interface. Output from Stage 3 included a heat map on the most promising aspects (red indicates more votes, n=5) with qualitative feedback highlights for additional considerations. The right pane illustrates an example of the prototype interface with a sample pop-out box. The numbers correspond to the flow of information and feedback from Stage 2 through to 3 and 4 using the first example (#1 in pink) to further illustrate flow with the dashed arrow.



STAGE 5: Usability Assessment – Results

Stage 5 results address Objective 3: Facilitate confidence in design decisions and empower user-centered technology development (Goal D).

Goal D: Evaluate a User-Driven, Practice-Relevant Prototype

Subjective usability was rated as "acceptable" with an average SUS score of 89.2, with the lowest and highest SUS scores of 72.5 and 97.5, respectively, translating to a B+ on the grade scale [52,53]. Mapping these scores onto the adjective ratings as described by Bangor et al, the majority of usability scores (5 out of 6) therefore fell between "excellent" and "best imaginable." In line with these quantitative results, users commented that, "It's quite intuitive, the key things were easily found," "It's a lot but it's easy to learn and it's colourful," "I'm not technologically inclined, but most things I was able to do intuitively," and "I think someone could use this if they were just thrown onto the floor with it."

As highlighted in Table 1, performance was rated comparably, with an average score of 16.8 and 15.2 for the AFINI-T and current systems, respectively. In the case of mental demand, time demand, and effort and frustration, subjective workload ratings were significantly lower for the AFINI-T system than the current system ($P<.05$). These results suggest the AFINI-T system is perceived to require less effort and lower overall workload than the current system. This is consistent with comments from the participants including the following: "[This would take a] huge burden off me as a clinician. This is hugely better than paper… there are no guestimates… I don't have to do work." and "It makes life so much easier."

For the AFINI-T system prototype in Stage 5, receptivity to the prototype was positive, with several areas identified for improvement. For example, the following was said regarding the general concept for the dietitian interface: "[It] would be good to personalize these specific needs and set it so the flags sent to nursing/PSW for these items based on what dietitian enters …This would save *a lot of time* especially if individualized."; "Capturing [supplement intake] would enable dietitians to monitor intervention adherence … If it shows up that they never have it, then great feedback to change the intervention."

A total of 2 technical experts completed a modified Ravden usability checklist evaluation with favorable ratings (Multimedia Appendix 3). Ratings across both raters for sections 1 to 8 were very satisfactory (7 out of 8 sections) or split among "satisfactory" and "very satisfactory" (1/8 sections) and mode for section 10 on system usability of "no problems." Consistent with comments from user testing, the main suggested area for improvement was to increase customizability options (eg, sort resident list in multiple ways, allow more flexibility in the order of operations such as allow charting before a picture is taken).

STAGE 6: Final Validation – Results

Receptivity of participants in Stage 6 was generally positive. The main reservation pertained to how the system would integrate with the current method and PointClickCare (corroborated in Stages 1, 5, and 6) and, more generally, the

workflow. For example, 3 webinar participants' direct messages were as follows: (1) "I love the idea of this system, we are concerned about workload, as well as if the systems (AFINI-T and PCC) talk to each other"; (2) "Would this be a separate system that would be linked to PCC?"; and (3) "I hope a PCC progress note is generated from any notes [a registered dietitian] adds."

Finally, participants expressed reservations regarding the proposed AFNI-T system. One dietitian expressed concern about overemphasizing the importance of nutrition "in a population that should have the main focus of just making sure [residents] are enjoying the food we are serving." There was also concern over how this will translate to Ontario Ministry of Health and Long-Term Care (MOHLTC) inspectors' inspections and the perception that using a system like this will take more time. In addition, it was stated that there was no perceived value to having access to more detailed nutrient data in the LTC population as, to them, the largest issue contributing to malnutrition is the impact dementia has on the calories consumed. However, they did suggest that if there was an ability to screen for residents to focus on only those at greater risk for malnutrition that the AFINI-T system would be helpful while still meeting the MOHLTC standards, as only those at risk for malnutrition are mandated to track food and fluid intake. This provides an interesting complementary perspective and warrants further probing and discussion.

Discussion

Summary

The overall purpose of this study was to investigate the gap for user acceptance studies and work toward a feasible food and fluid intake tracking solution for use in LTC through a participatory iterative design process and the creation and evaluation of a Goldilocks quality horizontal prototype. Specific contributions of this study were the following: (1) identify practice-relevant problems and solutions through user-centered participatory design, (2) remove feasibility-related barriers to uptake, and (3) facilitate confidence in design decisions and empower user-centered technology development.

We applied a rapid prototyping methodology via a modified Sprint process [22,23]. For the AFINI-T prototype, the data collection and design part of our modified sprint took place over 6 weeks rather than the suggested 5 days. This was because of the infeasibility of having an entire team of project advisors dedicated full time based on volunteered time, in addition to project advisors' regular full-time responsibilities. The discussion below is meant to elucidate several challenges in applying this framework in the academic research environment. In addition, we deepen our reflection on feedback received on the perception of the necessity of nutrient intake tracking in LTC with particular emphasis on this need within the dementia context.

Challenges of Applying the SPRINT Framework in Academic Research

Potential Challenges Around Organizing Activities

We were fortunate to have had our proposed workshop (Stage 1) accepted by the RIA and the Schlegel Villages as part of their annual Innovation Summit. This enabled us to gain momentum and build rapport from the in-person meeting and enabled many perspectives across several homes (within the same organization) to guide the direction for this project. If this infrastructure were not in place, coordinating the initial workshop would have been more challenging but not impossible with the following modifications. Initial discussion could have taken place with key stakeholders at targeted meetings (eg, quarterly dietitian meeting, and monthly team meetings). This would have required more travel and more time at the outset. The authors were also fortunate to have experience conducting applied research in the LTC environment. For others who may be newer to this approach, we recommend arranging a multiple day observation or volunteer experience to learn what the work environment is like to authentically understand the nuances of the needs and environment. We believe one key factor is to identify a necessary but highly inefficient and unreliable process.

Addressing the Need to Connect From a Distance

Many of the SPRINT activities were designed to be conducted in person. This was infeasible, given the time, distance, and multiple location constraints of project advisors' participation. As a result, many activities required modifications to approximate the intended function of the original activities. For example, the voting exercise and generating heat maps in Stage 2 were meant to be conducted in person with a group discussion. We made modifications by using the Qualtrics system for creating a Web-based survey paired with a Zoom meeting to enable discussion and screen sharing between each advisor and the lead author. In addition, tutorials needed to be developed and built into the Web-based survey (eg, how to make a vote and practice voting). It was crucial that this data collection tool development go through more than one iteration. We worked with an advisor from the support office to ensure the survey made sense, used sensitive language, and was streamlined enough to reduce potential frustration with completion.

Lessons Learned From Conducting Activities

Although Stages 1 to 6 all informed the design process, 1 specific opportunity for further enhancement was at Stage 6. We conducted a hybrid webinar survey to connect during a quarterly dietitian meeting. The concept of the AFINI-T system was completely new to the majority of participants, which made it difficult to build rapport with this group. However, we believe that at this stage of the design process, this was a strength; this may have helped participants provide candid, objective feedback. That said, there were several examples of difficulty in keeping webinar participants engaged. For example, the webinar was run with a brief adjournment for completion of a survey that was then used to encourage group discussion. The ability to take a poll during the webinar may have been more effective at keeping engagement. In addition, the method by which participants attended was inconsistent across locations.

For example, most participants joined individually; however, at venues where multiple participants joined from 1 location (eg, RD, director, and assistant director of food services), they filled out the corresponding survey together as well. This may have resulted in bias in some of the feedback collected but also enabled conversation and collaborative thought. Given the exploratory, qualitative nature of the feedback received during this stage, it does not undermine the results of previous stages and, for Stage 6, may have resulted in more critical appraisal from potential group discussion.

Timeliness in the Time-Constrained Dementia Care Context

One substantial difference between previous work on developing technology for consumer-centered nutrient intake tracking [17-20] and the work presented in this paper is that the purpose of our technology is to support tracking in a regulated LTC environment. This means considerations regarding consumer uptake and use are different than with general consumer market. For example, the novelty does not arise from tracking food and fluid intake per se; this is something that is already mandated for at-risk residents. Instead, the novelty is in improving the method for tracking beyond the current system in place. Other research involving diet tracking apps tends to focus on weight loss and is meant for tracking of an individual's food intake by the individual. Here, we seek to leverage LTC as an infrastructure already in place to conduct more efficient mandated multiperson monitoring.

The role of nutrition as part of a holistic care plan for individuals living with dementia is discussed in the 2015 European Society for Parenteral and Enteral Nutrition guidelines. They indicate that malnutrition contributes to disease progression and increased caregiver burden and that "nonpharmacological strategies like nutritional interventions are of particular interest as part of disease management" [54]. There is evidence to suggest that adhering to a particular pattern of dietary intake (eg, the Mediterranean diet) is associated with reduced cognitive decline [55]; however, these authors state "more conclusive evidence is needed to reach more targeted and detailed guidelines to prevent or postpone cognitive decline." Leveraging the necessity to monitor at-risk residents living in LTC through a novel, objective approach to food intake tracking may be beneficial for gaining new insights for defining guidelines.

Specifically considering the dementia care context and nutrition's role in the process, according to a 2016 systematic review [56], relatively few interventions have been conducted to explore the effect of food intake in mild cognitive impairment or dementia. They conclude that all 43 controlled interventions were at risk of bias and resulted in no consistent evidence either in support or against the effectiveness of nutrition-focused interventions [56]. By providing an alternative method for tracking, we seek to improve upon how these allocated resources are used and aim to provide more informative data. One future direction of the AFINI-T system is to use artificial intelligence to learn food preferences. Circling back to feedback we received in Stage 6, we wish to clarify that through this approach, the AFINI-T system may support caregivers' efforts in promoting enjoyment of food consumed for residents with communication changes as part of living the dementia journey. Within the scientific community context, in addition, the proposed AFINI-T system may enable knowledge discovery through a thorough automated approach to understanding dietary patterns in the LTC context and beyond.

Limitations

Between workshop participants and project advisors, 27 unique collaborators representing 15 different roles were engaged in this participatory iterative design process. This sample size is consistent with recent analogous health care–related, user-centered design as well as usability and feasibility studies [57-65], with sample sizes ranging from 5, as in the study by Khan et al [61], to 32, as in the study by Roberts et al [65]. Between 11 and 13 additional participants were involved in the webinar exercise and contributed to 9 survey responses (several individuals filled out a response together). Therefore, the total sample size ranged between 35 and 40; however, not all collaborators contributed to every aspect of the process (eg, user testing in Stage 5 comprised a subsample of 6 individuals). Although this sample size is consistent with early pilot-project prototyping [26,57-65], generalizability remains unclear. As the team of project advisors was relatively small and from the same organization, it will be important for the final product to be tested with a larger sample of users to make sure that the concepts captured more broadly generalize well to users' needs.

In terms of the physical design requirements, additional discussion is required, as the exact location to house the system remains unclear, as do size restrictions. What was gleaned, however, is that the AFINI-T system must work on the iPad, as this is what is currently in use. The acceptable level of accuracy target was not well defined with project advisors. That said, we can turn to the literature for some insight and important context. There is a tendency for frequent overestimation of food consumption [14,16]; in terms of degree of inaccuracy, estimates of food intake are typically over 50% for food items [16,66], with reported overestimation of food 22% of the time [14]. Furthermore, the source of error is said to be random [66], implying compensation is not possible with current methods. With the AFINI-T system, we should set our targets to be much more stringent, as the automated image-based system removes subjectivity. Careful documentation and exploration of the conditions where the system does not perform optimally will be necessary. One challenging situation is plates where the food items get mixed up over the course of the meal. However, even more crude estimates, where we assume equal eating distributions across types of foods for a plate average, would still improve on the current system as it eliminates subjectivity and reflects relative changes in mass and volume. In terms of time requirements and concerns raised in Stage 6, this is valid and is a next step. When the fully functional prototype is developed, it will be important to evaluate task completion time. Even if the AFINI-T system requires a comparable amount of time, it will yield a trove of powerful nutritional insights so direct comparison of approaches may be more complex than a simple timed trial.

Although it was clear that the project advisors were relatively diverse, no demographic information was collected; this should

be considered moving forward, especially when recruiting for a larger sample for user testing. A larger sample size for the final prototype will help deepen our understanding of usability. Finally, given the stage of this research, qualitative analyses were limited to extracting overarching themes across sources; an additional avenue for future work, pending completion of a high-fidelity prototype, is to conduct a more thorough qualitative analysis vetted in an evaluation framework (eg, grounded theory or narrative content analysis) alongside prototype testing and evaluation.

Conclusions

The purpose of this research was to conduct a multistage participatory iterative design sprint of a Goldilocks quality horizontal prototype for the AFINI-T system. Through input from 38 unique collaborators representing 15 distinct roles, design decisions were informed through the application of this user-centered participatory iterative design sprint. Output from these various stages suggest that although careful consideration for integration with the PointClickCare system is needed, as well as, more generally, policy expectations, project advisors are keen to try a technology like this. Advisors seem to be engaging with the AFINI-T prototype, are receptive to the idea, and are enjoying it. This modified participatory iterative design sprint was effective at understanding the problem space, making informed design decisions, and evaluating receptivity to a novel prototype, all within a compressed period of time (ie, 6 weeks). Next steps for the AFINI-T system include incorporation of learnings from this process and the development of a fully working prototype for additional user testing. We recommend this approach to others for general technology development.

Acknowledgments

The authors would like to acknowledge Amy Matharu and Susan Brown from the Schlegel-UW Research Institute for Aging for their assistance in coordinating this project within the Schlegel Villages, Jill Estioko and Heather Keller for their ongoing advocacy for and input into this project, Audrey Chung and Robert Amelard for their assistance with running the workshop, and, finally, all project advisors for their time and input into design decisions. This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs Program. This study has received ethics clearance by the University of Waterloo's Office of Research Ethics Board (#23124).

Authors' Contributions

KP conducted the design process, completed the analysis, and wrote this paper in full. JB and AW edited the paper. JB oversaw this project and provided ongoing feedback. AW is KP's supervisor.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of Stages 2 and 3 including purpose, methods, and results.

[PDF File (Adobe PDF File), 46KB - humanfactors_v6i2e13017_app1.pdf]

Multimedia Appendix 2

Summary of key inspiration concepts from commercially available online healthcare tools. Numbers in brackets correspond to corresponding design decisions.

[PDF File (Adobe PDF File), 24KB - humanfactors_v6i2e13017_app2.pdf]

Multimedia Appendix 3

A summary of the Ravden usability checklist evaluation conducted by 2 technical experts; section 9 was removed as it was not applicable to this version of the prototype.

[PDF File (Adobe PDF File), 23KB - humanfactors_v6i2e13017_app3.pdf]

References

1. Keller HH, Østbye T, Goy R. Nutritional risk predicts quality of life in elderly community-living Canadians. J Gerontol A Biol Sci Med Sci 2004 Jan;59(1):68-74. [Medline: 14718488]
2. Pirlich M, Lochs H. Nutrition in the elderly. Best Pract Res Clin Gastroenterol 2001 Dec;15(6):869-884. [doi: 10.1053/bega.2001.0246] [Medline: 11866482]

3. Thomas DR, Ashmen W, Morley JE, Evans WJ. Nutritional management in long-term care: development of a clinical guideline. Council for Nutritional Strategies in Long-Term Care. J Gerontol A Biol Sci Med Sci 2000 Dec;55(12):M725-M734. [Medline: 11129394]
4. Goates S, Du K, Braunschweig CA, Arensberg MB. Economic burden of disease-associated malnutrition at the state level. PLoS One 2016;11(9):e0161833 [FREE Full text] [doi: 10.1371/journal.pone.0161833] [Medline: 27655372]
5. Russell C. The impact of malnutrition on healthcare costs and economic considerations for the use of oral nutritional supplements. Clin Nutr Suppl 2007 Jan;2(1):25-32. [doi: 10.1016/j.clnu.2007.04.002]
6. Brownie S. Why are elderly individuals at risk of nutritional deficiency? Int J Nurs Pract 2006 Apr;12(2):110-118. [doi: 10.1111/j.1440-172X.2006.00557.x] [Medline: 16529597]
7. Ontario Long Term Care Association. 2016. This is long-term care 2016 URL: https://www.oltca.com/oltca/documents/reports/tiltc2016.pdf [accessed 2019-04-18] [WebCite Cache ID 77iaC2Lnr]
8. Harris-Kojetin L, Sengupta M, Park-Lee E, Valverde R. Long-term care services in the United States: 2013 overview. Vital Health Stat 3 2013 Dec(37):1-107 [FREE Full text] [Medline: 26158640]
9. Keller H, Carrier N, Slaughter S, Lengyel C, Steele C, Duizer L, et al. Prevalence and determinants of poor food intake of residents living in long-term care. J Am Med Dir Assoc 2017 Nov 01;18(11):941-947. [doi: 10.1016/j.jamda.2017.05.003] [Medline: 28668663]
10. Bell CL, Tamura BK, Masaki KH, Amella EJ. Prevalence and measures of nutritional compromise among nursing home patients: weight loss, low body mass index, malnutrition, and feeding dependency, a systematic review of the literature. J Am Med Dir Assoc 2013 Feb;14(2):94-100. [doi: 10.1016/j.jamda.2012.10.012] [Medline: 23246236]
11. Dietitians of Canada. 2013 Apr. Best Practices for Nutrition, Food Service and Dining in Long Term Care Homes Internet URL: https://www.dietitians.ca/Downloads/Public/2013-Best-Practices-for-Nutrition,-Food-Service-an.aspx [accessed 2019-04-16] [WebCite Cache ID 77g6rwzxI]
12. Sloane P, Ivey J, Helton M, Barrick A, Cerna A. Nutritional issues in long-term care. J Am Med Dir Assoc 2008 Sep;9(7):476-485. [doi: 10.1016/j.jamda.2008.03.005] [Medline: 18755420]
13. Andrews Y, Castellanos V. Development of a method for estimation of food and fluid intakes by nursing assistants in long-term care facilities: a pilot study. J Am Diet Assoc 2003 Jul;103(7):873-877. [doi: 10.1053/jada.2003.50168] [Medline: 12830027]
14. Simmons S, Reuben D. Nutritional intake monitoring for nursing home residents: a comparison of staff documentation, direct observation, and photography methods. J Am Geriatr Soc 2000 Feb;48(2):209-213. [Medline: 10682952]
15. Simmons S, Schnelle J. Feeding assistance needs of long-stay nursing home residents and staff time to provide care. J Am Geriatr Soc 2006 Jun;54(6):919-924. [doi: 10.1111/j.1532-5415.2006.00812.x] [Medline: 16776786]
16. Castellanos V, Andrews Y. Inherent flaws in a method of estimating meal intake commonly used in long-term-care facilities. J Am Diet Assoc 2002 Jun;102(6):826-830. [Medline: 12067049]
17. Kong F. Michigan Technological University. 2012. Automatic food intake assessment using camera phones URL: https://digitalcommons.mtu.edu/cgi/viewcontent.cgi?article=1493&context=etds
18. Meyers A, Johnston N, Rathod V, Korattikara A, Gorban A, Silberman N, et al. Proceedings of the IEEE International Conference on Computer Vision. Im2Calories: towards an automated mobile vision food diary URL: http://openaccess.thecvf.com/content_iccv_2015/html/Meyers_Im2Calories_Towards_an_ICCV_2015_paper.html [accessed 2019-04-16] [WebCite Cache ID 77g7aXflq]
19. Okamoto K, Yanai K. An automatic calorie estimation system of food images on a smartphone. In: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. 2016 Presented at: MADiMa '16; October 16, 2016; Amsterdam, The Netherlands p. 63-70 URL: https://dl.acm.org/citation.cfm?id=2986040 [doi: 10.1145/2986035.2986040]
20. Pouladzadeh P, Shirmohammadi S, Yassine A. You are what you eat: so measure what you eat!. IEEE Instrum Meas Mag 2016 Feb;19(1):9-15. [doi: 10.1109/MIM.2016.7384954]
21. Parent M, Niezgoda H, Keller H, Chambers L, Daly S. Comparison of visual estimation methods for regular and modified textures: real-time vs digital imaging. J Acad Nutr Diet 2012 Oct;112(10):1636-1641. [doi: 10.1016/j.jand.2012.06.367] [Medline: 23017574]
22. Knapp J, Zeratsky J, Kowitz B. Sprint: How To Solve Big Problems And Test New Ideas In Just Five Days. New York: Simon & Schuster; 2019.
23. Norman D. The Design Of Everyday Things: Revised And Expanded Edition. New York: Basic Books; 2019.
24. Nielsen J. Usability Engineering. Cambridge: Morgan Kaufmann; 1994.
25. Boger J, Jackson P, Mulvenna M, Sixsmith J, Sixsmith A, Mihailidis A, et al. Principles for fostering the transdisciplinary development of assistive technologies. Disabil Rehabil Assist Technol 2017 Dec;12(5):480-490. [doi: 10.3109/17483107.2016.1151953] [Medline: 27052793]
26. Carr E, Babione J, Marshall D. Translating research into practice through user-centered design: an application for osteoarthritis healthcare planning. Int J Med Inform 2017 Dec;104:31-37. [doi: 10.1016/j.ijmedinf.2017.05.007] [Medline: 28599814]
27. Postma CE, Zwartkruis-Pelgrim E, Daemen E, Du J. Challenges of doing empathic design: experiences from industry. Int J Des 2012;6 [FREE Full text]

28. Sanders L. On Modeling: an evolving map of design practice and design research. Interactions 2008;15(6):13-17 [FREE Full text] [doi: 10.1145/1409040.1409043]
29. Apple Inc. Apple Inc. Human Interface Guidelines: Navigation Bars URL: https://developer.apple.com/design/human-interface-guidelines/ios/bars/navigation-bars/ [accessed 2019-04-18] [WebCite Cache ID 77iayk6FA]
30. Babich N. UX Planet. Button UX Design: Best Practices, Types and States Internet URL: https://uxplanet.org/button-ux-design-best-practices-types-and-states-647cf4ae0fc6?gi=662ec0e75fe1 [accessed 2019-04-18] [WebCite Cache ID 77ib1Bfip]
31. UX Planet. 7 rules for mobile UI button design Internet URL: https://uxplanet.org/7-rules-for-mobile-ui-button-design-e9cf2ea54556?gi=8b07e8118285 [accessed 2019-04-18] [WebCite Cache ID 77ib4Rhwc]
32. Nielsen Norman Group. 1995. 10 usability heuristics for user interface design URL: https://medium.com/@toddohanian/10-usability-heuristics-for-user-interfaces-in-web-design-c179aa39b54e [accessed 2019-04-16] [WebCite Cache ID 77g9kmtVX]
33. Shneiderman B, Plaisant C, Cohen M, Jacobs S, Elmqvist N, Diakopoulos N. Designing The User Interface: Strategies For Effective Human-computer Interaction (6th Edition). New York: Pearson; 2019.
34. Brooke J. SUS-A quick and dirty usability scale. In: Usability evaluation in industry. London: Taylor & Francis; 1996.
35. Ravden S, Johnson G. Evaluating usability of human-computer interfaces: a practical method. New York: Halsted Press; 1989.
36. Hart S, Staveland L. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Advances in Psychology 1988;52:139-183 [FREE Full text] [doi: 10.1016/S0166-4115(08)62386-9]
37. Corritore C, Kracher B, Wiedenbeck S. On-line trust: concepts, evolving themes, a model. Int J Hum-Comput Stud 2003 Jun;58(6):737-758. [doi: 10.1016/S1071-5819(03)00041-7]
38. Fogg B, Tseng H. The elements of computer credibility. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 1999 Presented at: CHI '99; May 15-20, 1999; Pittsburgh, Pennsylvania, USA p. 80-87 URL: https://dl.acm.org/errorpgs/403.html [doi: 10.1145/302979.303001]
39. YouTube. 2016. Aprima Overview URL: https://www.youtube.com/watch?v=Q0d0Oa-xYH4 [accessed 2019-04-16] [WebCite Cache ID 77gE5Q4zU]
40. YouTube. 2017. PrognoCIS EHR Software Demo Video URL: https://www.youtube.com/watch?v=phzFiyq6d8Q [accessed 2019-04-16] [WebCite Cache ID 77gFt8Jpz]
41. YouTube. 2015. ChiroSpring Overview Video - Cloud-based Chiropractic Practice Management Software URL: https://www.youtube.com/watch?v=3Z6pAXhAHOc [accessed 2019-04-16] [WebCite Cache ID 77gFz7vjp]
42. Capterra. Best Electronic Medical Records (EMR) Software URL: https://www.capterra.com/electronic-medical-records-software/ [accessed 2018-11-26] [WebCite Cache ID queryurlhttps3A2F2Fwwwcapterracom2Felectronicmedicalrecordssoftwareampdate20181126]
43. Software Advice. Top Electronic Medical Records Software - 2018 Reviews URL: https://www.softwareadvice.com/ca/medical/electronic-medical-record-software-comparison/ [accessed 2019-04-18] [WebCite Cache ID 77ibsSSOq]
44. Hart SG. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2006 Oct 01. Nasa-Task Load Index (NASA-TLX); 20 Years Later URL: https://journals.sagepub.com/doi/pdf/10.1177/154193120605000909 [accessed 2019-04-16] [WebCite Cache ID 77gB1W0Z2]
45. Stanton N, Salmon P, Rafferty L. Human factors methods: a practical guide for engineering and design. Burlington: Ashgate Publishing, Ltd; 2013.
46. Luximon A, Goonetilleke R. Simplified subjective workload assessment technique. Ergonomics 2001 Feb 20;44(3):229-243. [doi: 10.1080/00140130010000901] [Medline: 11219757]
47. Vidulich MA, Tsang PS. Techniques of subjective workload assessment: a comparison of SWAT and the NASA-Bipolar methods. Ergonomics 1986 Nov;29(11):1385-1398. [doi: 10.1080/00140138608967253]
48. Rhemtulla M, Brosseau-Liard P, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol Methods 2012 Sep;17(3):354-373. [doi: 10.1037/a0029315] [Medline: 22799625]
49. Stanton NA, Young MS. Guide to Methodology in Ergonomics: Designing for Human Use. Boca Raton: CRC Press; 2002.
50. Field A. Comparing two means. In: Discovering Statistics Using Ibm Spss Statistics, 4th Edition. Thousand Oaks: Sage Publications Ltd; 2019:364-370.
51. Norman G, Streiner DL. Comparing two groups. In: Biostatistics the Bare Essentials. New York: McGraw-Hill Education; 2008:70-73.
52. 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.
53. Bangor A, Kortum P, Miller J. An empirical evaluation of the system usability scale. Int J Hum Comput Stud 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
54. Volkert D, Chourdakis M, Faxen-Irving G, Frühwald T, Landi F, Suominen M, et al. ESPEN guidelines on nutrition in dementia. Clin Nutr 2015 Dec;34(6):1052-1073. [doi: 10.1016/j.clnu.2015.09.004] [Medline: 26522922]

55. van de Rest O, Berendsen A, Haveman-Nies A, de Groot LC. Dietary patterns, cognitive decline, and dementia: a systematic review. Adv Nutr 2015 Mar;6(2):154-168 [FREE Full text] [doi: 10.3945/an.114.007617] [Medline: 25770254]
56. Abdelhamid A, Bunn D, Copley M, Cowap V, Dickinson A, Gray L, et al. Effectiveness of interventions to directly support food and drink intake in people with dementia: a systematic review and meta-analysis. BMC Geriatr 2016 Jan 22;16:26 [FREE Full text] [doi: 10.1186/s12877-016-0196-3] [Medline: 26801619]
57. Diamantidis C, Ginsberg J, Yoffe M, Lucas L, Prakash D, Aggarwal S, et al. Remote usability testing and satisfaction with a mobile health medication inquiry system in CKD. Clin J Am Soc Nephrol 2015 Aug 07;10(8):1364-1370 [FREE Full text] [doi: 10.2215/CJN.12591214] [Medline: 26220816]
58. Gray CS, Gill A, Khan A, Hans P, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
59. Hohenstein J, O'Dell D, Murnane E, Lu Z, Erickson D, Gay G. Enhancing the usability of an optical reader system to support point-of-care rapid diagnostic testing: an iterative design approach. JMIR Hum Factors 2017 Nov 21;4(4):e29 [FREE Full text] [doi: 10.2196/humanfactors.8621] [Medline: 29162559]
60. Jahn M, Porter B, Patel H, Zillich A, Simon S, Russ A. Usability assessment of secure messaging for clinical document sharing between health care providers and patients. Appl Clin Inform 2018 Apr;9(2):467-477 [FREE Full text] [doi: 10.1055/s-0038-1660521] [Medline: 29949815]
61. Khan S, McCullagh L, Press A, Kharche M, Schachter A, Pardo S, et al. Formative assessment and design of a complex clinical decision support tool for pulmonary embolism. Evid Based Med 2016 Feb;21(1):7-13. [doi: 10.1136/ebmed-2015-110214] [Medline: 26718820]
62. Kushniruk A, Senathirajah Y, Borycki E. Towards a usability and error. Stud Health Technol Inform 2017;245:763-767. [Medline: 29295201]
63. Nagykaldi Z, Jordan M, Quitoriano J, Ciro C, Mold J. User-centered design and usability testing of an innovative health-related quality of life module. Appl Clin Inform 2014;5(4):958-970 [FREE Full text] [doi: 10.4338/ACI-2014-08-RA-0067] [Medline: 25589910]
64. Rajan J, Moura J, Gourley G, Kiso K, Sizilio A, Cortez A, et al. Understanding the barriers to successful adoption and use of a mobile health information system in a community health center in São Paulo, Brazil: a cohort study. BMC Med Inform Decis Mak 2016 Dec 17;16(1):146 [FREE Full text] [doi: 10.1186/s12911-016-0385-1] [Medline: 27855685]
65. Roberts S, Marshall A, Gonzalez R, Chaboyer W. Technology to engage hospitalised patients in their nutrition care: a qualitative study of usability and patient perceptions of an electronic foodservice system. J Hum Nutr Diet 2017 Dec;30(5):563-573. [doi: 10.1111/jhn.12467] [Medline: 28211190]
66. Bingham S. Limitations of the various methods for collecting dietary intake data. Ann Nutr Metab 1991;35(3):117-127. [doi: 10.1159/000177635] [Medline: 1952811]

Abbreviations

AFINI-T: Automated Food Imaging and Nutrient Intake Tracking
LTC: long-term care
MOHLTC: Ministry of Health and Long-Term Care
PSW: personal support worker
RD: registered dietitian
RIA: Research Institute for Aging
RTLX: Raw Task Load Index
SUS: Subjective Usability Scale

Edited by G Eysenbach; submitted 03.12.18; peer-reviewed by C Berridge, B Chaudhry; comments to author 07.02.19; revised version received 27.02.19; accepted 06.04.19; published 09.05.19.

Please cite as:
Pfisterer KJ, Boger J, Wong A
Prototyping the Automated Food Imaging and Nutrient Intake Tracking System: Modified Participatory Iterative Design Sprint
JMIR Hum Factors 2019;6(2):e13017
URL: http://humanfactors.jmir.org/2019/2/e13017/
doi:10.2196/13017
PMID:31094336

Original Paper

Anthropomorphism of Robots: Study of Appearance and Agency

Charles R Crowell[1], PhD; Jason C Deska[2], PhD; Michael Villano[1], PhD; Julaine Zenk[1], MA; John T Roddy Jr[1], BA

[1]Department of Psychology, University of Notre Dame, Notre Dame, IN, United States

[2]Department of Psychology, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Charles R Crowell, PhD
Department of Psychology
University of Notre Dame
462 Corbett Family Hall
Notre Dame, IN, 46556
United States
Phone: 1 574 276 8581
Fax: 1 574 271 2058
Email: ccrowell@nd.edu

Abstract

Background: As the prevalence of robots increases each year, understanding how we anthropomorphize and interact with them is extremely important. The three-factor theory of anthropomorphism, called the Sociality, Effectance, Elicited agent Knowledge model, guided this study. As anthropomorphism involves a person making attributions of human likeness toward a nonhuman object, this model implies that anthropomorphism can be influenced either by factors related to the person or the object.

Objective: The aim of this study was to explore factors influencing the anthropomorphism of robots, specifically the robot's appearance (humanoid vs nonhumanoid) and agency (autonomous vs nonautonomous). We expected a humanoid robot would be anthropomorphized to a greater extent than one that was nonhumanoid. In addition, we expected that inducing an agency belief to the effect that a robot was making its own decisions would increase anthropomorphism compared with a nonagency belief that the robot was being remotely controlled by a human. We also sought to identify any role gender might play in anthropomorphizing the robot.

Methods: Participants (N=99) were primed for agency or nonagency belief conditions and then saw a brief video depicting either a humanoid or nonhumanoid robot interacting with a confederate. After viewing the video, they completed 4 measures: perception to humanoid robots scale (PERNOD), the Epley anthropomorphic adjectives measure, the Fussel anthropomorphic adjective checklist, and the Anthropomorphic Tendencies Scale (ATS).

Results: Findings with the PERNOD scale indicated subjects did perceive the 2 robots differently, $F_{6,86}$=6.59, P<.001, which means the appearance manipulation was effective. Results with the Epley adjectives indicated that participants were more willing to attribute humanlike behavioral traits to the nonhumanoid rather than the humanoid robot, $F_{1,91}$=5.76, P=.02. The Fussel adjective checklist results showed that subjects were more willing to attribute humanlike social qualities to the remote controlled than the autonomous robot, $F_{1,91}$=5.30, P=.02. Finally, the ATS revealed the only gender effects in this study, with females reporting more endorsement of anthropomorphism for pets (P=.02) and less for showing negative emotions toward anthropomorphized objects (P<.001) if they had witnessed the humanoid rather than the nonhumanoid robot.

Conclusions: Contrary to our expectations, participants were less willing to make humanlike attributions toward a robot when its morphology was more humanlike and were more willing to make those attributions when they were told that the robot was being remotely controlled by a person rather than acting on its own. In retrospect, these outcomes may have occurred because the humanoid robot used here had a smaller overall stature than the nonhumanoid robot, perhaps making it seem more toylike and because subjects made attributions toward the person behind the remote-controlled robot rather than toward the robot itself.

(JMIR Hum Factors 2019;6(2):e12629) doi:10.2196/12629

KEYWORDS

psychology, social; social perception; theory of mind; cognitive science; perception; cognition; robotics; telerobotics; human factors engineering

Introduction

Anthropomorphism and the Sociality, Effectance, Elicited Agent Knowledge Model

Anthropomorphism can be defined as "the tendency to imbue the real or imagined behavior of nonhuman agents with human like characteristics, motivations, intentions, or emotions" [1]. Instances of anthropomorphism occur all around us on a daily basis, from the tendency to imbue pets with human like traits [2] to the attribution of human like characteristics to deities [1] and even to personal computers [3]. As computers and robots become increasingly ubiquitous, our understanding of how we anthropomorphize robots in human robot interactions (HRI) will become more important as well [4].

The Sociality, Effectance, Elicited Agent Knowledge Model of Anthropomorphism

Epley et al [1,2] have proposed a 3-factor theory, called the Sociality, Effectance, Elicited agent Knowledge (SEEK) model, to explain why human beings anthropomorphize. The first factor in the SEEK model is *sociality motivation*, or the human desire for social connections [1]. Humans are social animals with a strong need to establish and maintain a sense of interpersonal connection to others. Sociality motivation increases the tendency to search actively for sources of social connections in one's environment or to invent those connections when necessary.

The second SEEK factor is *effectance motivation*, or the need to understand, control, and interact effectively with the environment [1]. This factor can give rise to the desire to understand the behavior of nonhuman agents by projecting onto them more familiar human traits. In this way, anthropomorphism serves as a tool to facilitate understanding of (and potentially control over) unfamiliar agents by making them more humanlike. This tendency can be exacerbated in situations where the behavior of a nonhuman agent cannot be accounted for easily by other explanations. For example, researchers have reported the results of a study in which participants were shown a brief video clip of 2 dogs interacting with each other where 1 dog was more behaviorally unpredictable [2]. Results indicated that participants were more likely to ascribe human like qualities to the less predictable dog. This finding seems consistent with the possibility that attributions of human like agency to nonhuman entities, including robots, are more likely when alternative accounts of agent behavior are not readily available [5].

The third SEEK factor, *elicited agent knowledge*, refers to the extent to which people have, and can, apply relevant anthropocentric knowledge to objects or entities that might be targets for the attribution of human like qualities. Homocentric knowledge often serves as the basis for making inferences about lesser known, nonhuman agents [1]. It follows, then, that physical appearance and movement of a nonhuman agent might be important factors in guiding elicited agent knowledge. That is, the evocation and application of anthropocentric knowledge might be facilitated by the morphological and kinetic similarity between human and nonhuman agents [1,4,6]. Recent studies with robots consistent with this possibility has shown that greater robot human likeness affects the receptivity of humans to advice provided by a robot [7], the extent to which humans will empathize with a robot [8], how much credit a human will take in a joint human robot task for successful task completion [9], and even the types of tasks for which a robot is deemed suitable [10,11]. Furthermore, when a person believes a robot shares his or her own gender, that person is more likely to attribute a human mind to the robot [12].

Gender Differences in Human Robot Interaction and Anthropomorphism

Another factor potentially influencing the tendency of humans to anthropomorphize robots is gender. Several studies have explored how gender affects HRI. One study showed that males and females provided significantly different answers to social desirability questions asked by a voice that was either disembodied or coming from a robot [13]. Females showed less social desirability scores with the disembodied voice, whereas males showed less social desirability with the robot. These findings suggest that males may have felt more open and honest with the robot than did females. In any case, these results indicate that males perceived the robot differently from females [13].

A study of proxemics, or the use of personal space and comfortable distances, involving robots examined personal preferences when a robot could approach a participant either directly from the front or at an angle from the side [14]. Results showed that although females were largely indifferent as to whether the robot approached from the front or side, males were much more uncomfortable when the robot approached from directly in front of them as opposed to the side [14]. Researchers suggested that a front approach may have been perceived as more combative by the males [14].

Other studies have looked more closely at opinions toward robots based on a person's gender. The Negative Attitudes toward Robots Scale [15] has been used numerous times to show that females tend to have significantly stronger negative opinions toward robots than males [16-18]. One study found females had lower rates of robot liking and higher rates of *Robotphobia* than their male counterparts [19]. Researchers have also surveyed adult opinions on a mechanical robot at a public mall and found that females found the robot unpredictable, whereas males found the robot helpful [20]. Looking closer at helpfulness, another study found that males, regardless of age, rated a health care robot as more useful and were more hopeful for its future development than were their female counterparts, both before and after interacting with a health care robot [21].

Furthermore, it has been demonstrated that even the tendency to anthropomorphize itself can be impacted by gender [22]. Using 2 scales directly measuring a person's tendency to anthropomorphize pets, gods, or artifacts, investigators found that females were more likely than males to anthropomorphize animals but found no differences in the tendency to anthropomorphize artifacts. However, in this study, the category of *artifacts* included both robots and mechanical devices, such as vehicles or computers, so no specific anthropomorphism rating for robots could be determined from this study.

Purpose of the Study

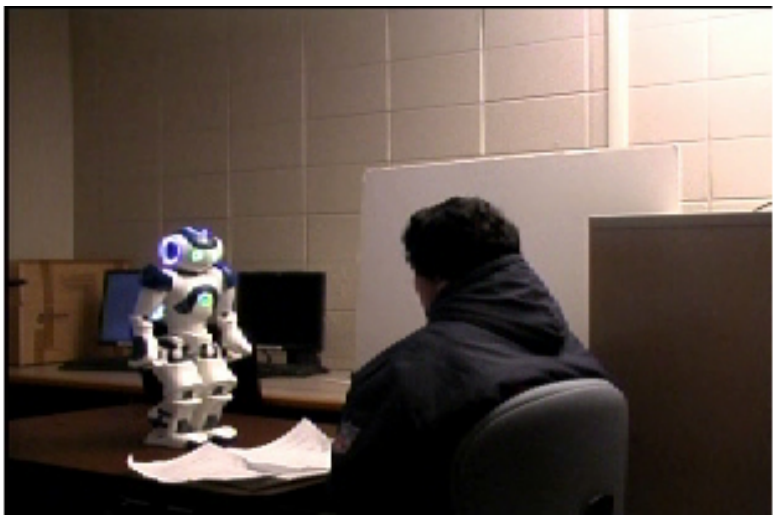
The SEEK model represents a psychological theory of the determinants of anthropomorphism, which may have broader applicability to our understanding of why people make attributions of human like qualities to diverse nonhuman entities, including machines and robots. Preliminary support for the applicability of the SEEK model to the anthropomorphism of robots has already been provided, particularly as it relates to sociality and effectance motivation [5,23,24], and, to a lesser extent, elicited agent knowledge [12]. The primary purpose of this study was to further examine the SEEK factor of elicited agent knowledge by evaluating 2 specific hypotheses related to that factor.

Hypothesis 1

Participants should engage in more anthropomorphism toward a humanoid robot than toward one that is nonhumanoid because greater similarity of appearance to a person should allow participants to bring more self-knowledge to bear on their understanding of and attributions toward the humanoid agent. To test this hypothesis, we employed an *appearance* manipulation involving 2 different robots: one robot having a distinctly humanoid form, whereas the other clearly having a much less human like appearance (see Figures 1 and 2).

Figure 1. A screen capture from the experiment video of the nonhumanoid robot in dialogue with confederate.

Figure 2. A screen capture from the experiment video of the humanoid robot in dialogue with confederate.



Hypothesis 2

Participants should anthropomorphize an autonomous robot more than a nonautonomous robot because of the greater ease with which autonomy allows humans to apply their own anthropocentric knowledge as a means of understanding the autonomous nonhuman agent's behavior. We tested this hypothesis by using an *agency* manipulation to induce different beliefs about either the humanoid or nonhumanoid robots. One belief was that both robots were fully autonomous and capable of acting on their own volition, whereas the second belief was that an experimenter in another room was controlling the robots. This *agency* manipulation was intended to provide participants with distinctively different explanations for the behavior of the robot to which they were exposed. We expected participants to identify more closely with the autonomous agent, regardless of its appearance.

As noted above, females have a more negative view of robots than males [15-17] and have a greater tendency than males to anthropomorphize animals [22]. However, the implications of

these findings for gender-based differences in anthropomorphism of robots are not clear. Thus, a secondary purpose of this study was to include gender as an additional factor in the evaluation of both hypotheses 1 and 2.

Methods

Participants

Participants consisted of 99 undergraduate students, 52 males and 47 females, between the ages of 18 and 22 years, enrolled at a midsized, private, Midwestern university. Participants voluntarily chose to be in this study and received course credit for their participation. We treated participants in accordance with the ethical standards of the American Psychological Association and the institutional review board approved the research protocol.

Design

The design of this study conformed to a 2 (appearance: humanoid vs nonhumanoid) × 2 (agency: autonomous vs nonautonomous) × 2 (participant gender: male vs female) factorial design, with all factors varied between subjects. Participants were assigned randomly to each of the 4 groups.

Materials

The following materials used in this study are organized according to whether they were administered before the experimental manipulations (pretest materials), during the manipulations themselves (experimental materials), or after the manipulations (posttest materials).

Pretest Materials

Several pretest scales were used to verify that our 4 groups did not differ on factors that could influence the results other than the explicitly manipulated factors. A total of 3 of the pretest scales, Desire for Control [25], University of California, Los Angeles (UCLA) Loneliness [26], and Need for Cognition [27] were used because they have been directly tied to the factors involved in the SEEK model [1,2]. Desire for Control and Loneliness are relevant to the SEEK factors of effectance motivation and sociality motivation, whereas Need for Cognition is linked to one's ability to employ elicited agent knowledge [1,2]. By comparing the groups on these scales, we could verify that our independent groups in this study were not different in these SEEK related factors before experiencing our experimental manipulations. A shortened version of the Marlowe-Crowne scale [28,29], a questionnaire measuring social desirability, was included as a pretest check that our groups also did not differ in social desirability, which could influence their responses to the questionnaires used in this study.

Experimental Materials

During the experimental phase of this study, each participant was exposed to 1 agency manipulation story (either autonomous or nonautonomous) and 1 robot interaction video (depicting either the humanoid or nonhumanoid robot). During the course of the experiment, participants made no direct contact with either robot. All experimental materials listed below can be found in Multimedia Appendix 1.

Agency Manipulation Stories

The agency manipulation (autonomous vs nonautonomous) used in this study was delivered via 1 of 2 different stories read to participants just before they saw their designated humanoid or nonhumanoid robot video (see Multimedia Appendix 1). Both videos depicted the 2 morphologically distinct robots doing exactly the same things. The autonomous story said that the robot had artificial intelligence and the capability to perform fully autonomous behavior. In contrast, for the nonautonomous condition, the story explained that a human controlled the robot from another room.

Appearance Manipulation Videos

On the basis of previous work using virtual reality environments indicating that human evaluations of virtual robots are comparable in many respects with those obtained from observing similar physically present robots [30,31], along with the work reported by another study showing that measures taken via live interactions with a robot are comparable with those from video based interactions [32], we believed that video exposure to robots would produce effects comparable with those obtained from direct, physical exposure. Therefore, our appearance manipulation involved having participants view a brief video of a confederate experimenter interacting with either the humanoid or a nonhumanoid robot, depending on the appearance condition to which participants were assigned (see Multimedia Appendices 2 and 3).

The nonhumanoid robot was a PeopleBot, obtained from MobileRobots Inc (see Figure 1 and Multimedia Appendix 3). The nonhumanoid robot was approximately 5 feet tall and had small speakers sitting on either side of the upper shelf under the cameras and slightly elongated grippers to provide the impression of arm like appendages. The humanoid robot was a Nao Academics Edition, Version 3 Plus obtained from Aldebaran Robotics (see Figure 2 and Multimedia Appendix 2), which was approximately 1.9 feet tall weighing approximately 9.5 lbs.

Drawing on previous work showing humans prefer to interact with telepresence robots at eye level [33], the robots in this study were oriented such that the tops of their heads were near the top of the camera frame, making the heads of both equidistant from the floor. This resulted in positioning the confederate's head and gaze at approximately the same viewing angle for both robots. To do this, the Nao robot stood on a table, whereas the PeopleBot remained on the floor, and the confederate remained seated in both videos (see Figures 1 and 2). Consistent with the findings of a longitudinal study by of HRI [34], we believed this arrangement would allow participants to respond more to robot appearance than to robot height. In addition, differences of up to 0.2 meters (0.7 ft) in robot height do not significantly influence opinions toward or comfort with robots [35,36].

The same script was used to create the 2 videos depicting identical interactions of a student with either the humanoid or nonhumanoid robot. The script depicts a conversation between the robot and a student in which the robot described and demonstrated some of its capabilities and then engaged the student in a brief discussion about college football. All robot

movements were carefully selected so as to be comparable between the 2 robot platforms, and the same voice was used for both robots. The entire video lasted approximately 3.5 min.

Posttest Materials

Our posttest measures involved published scales previously used to determine how subjects perceived and anthropomorphized the robot they saw in the video.

Perception of Humanoid Robots Scale

The perception of humanoid robots' scale, known as Perception of Humanoid Robots Scale (PERNOD) [37], was employed in this study as an appearance manipulation check to see how similar or different our participants viewed the humanoid and nonhumanoid robots we used. The PERNOD evaluates a participant's perception of a particular robot on 6 separate dimensions: *graceful* related to the quickness or slowness with which it moved; *expressive* related to how the robot communicates emotion or friendliness; *useful* related to potential utility of the robot for humans; *controllable* related to how subservient to humans it appears; *durable*, which reflects a lack of concern about fragility or breakability of the robot; and *smooth*, which refers to the look or physical appearance of the platform being not angular or coarse. A 7-point scale was used for all items, and the scoring was such that higher values indicate stronger alignment with a dimension.

Epley et al Anthropomorphic Items

One measure of anthropomorphism used in this study was derived from several items used in a study with pet owners [2]. These measures consisted of 7 anthropomorphic (thoughtful, considerate, sympathetic, embarrassable, creative, devious, and jealous) and 7 behavioral (aggressive, agile, active, energetic, fearful, lethargic, and muscular) trait adjectives that participants were asked to rank from 1 to 14 in order of decreasing applicability to the robot they saw in the video. Separate sums of ranks were computed for both anthropomorphic and behavioral trait adjective categories for each participant. The scores were then reverse coded such that a higher sum of ranks signified that the adjectives in that category were rated as more applicable to the robot. These 2 groups of anthropomorphic and behavioral traits were created by Epley et al, and we retained this categorization for this study [2].

Fussel Adjective Checklist

A second measure of anthropomorphism used in this study was based on an adjective checklist used in an earlier study [38]. This checklist consisted of 40 adjectives, 10 in each of 4 categories: human sociability, other human, robotic, and false fillers [38]. There were both positive and negative adjectives in both human categories. A total of 2 of the other human adjectives were gender related and were separated, whereas the remaining 8 referred to what can be called *human personality* traits. A third category pertained to characteristics of robots, which itself can be subdivided into characteristics that clearly were mechanical and those that were not. The fourth category consisted of characteristics very likely to be rated as false for both humans and robots but which could not be subdivided in any obvious way. Our revised breakdown of the original 40-item [38] adjective checklist is shown in Table 1. It is important to note here that the adjectives in this table are exactly the same as those employed in the reference study, only their organization has been changed to distinguish positive or negative and mechanical or nonmechanical subcategories. Subjects designated each adjective as either true or false for the robot (humanoid or nonhumanoid) they saw.

For all of the categories in Table 1, except Gender, we computed a proportion of true responses across the adjectives in that category for each subject. This resulted in separate proportions for each subject for the categories of human social positive, human social negative, human personality positive, human personality negative, robotic mechanical, robotic nonmechanical, and false fillers. The 2 gender categories were mutually exclusive such that participants assigned the 1 robot they saw either a male or a female designation.

Table 1. The reorganized Fussel adjective checklist.

Category	Adjectives
Human social positive	Friendly, polite, sensitive, caring, and sociable
Human social negative	Rude, obnoxious, cold, impatient, and aggressive
Human personality positive	Organized, thorough, curious, and persistent
Human personality negative	Nervous, distractible, shallow, and disorganized
Gender	Male and female
Robotic nonmechanical	Android, artificial, automaton, mechanical, controllable, and robotic
Robotic mechanical	Synthetic, breakable, software, and portable
False fillers	Animal, wooden, wet, smelly, tubular, ceramic, cotton, striped, roasted, and bloody

Anthropomorphic Tendencies Scale

A third measure of anthropomorphism was the Anthropomorphic Tendencies Scale (ATS) [39]. The ATS measured 4 subscale dimensions of anthropomorphism: extreme anthropomorphism (the attribution of human like qualities to physical objects such as backpacks or cars), anthropomorphism toward pets, anthropomorphism toward gods or deities, and negative anthropomorphism, which reflects a tendency to display negative emotions toward physical objects such as cars or computers. These dimensions were rated on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Mean scores that were greater than 4 represented agreement, means scores of 2 and below represented disagreement, whereas mean scores of 3

reflected a neutral rating. Although this scale may measure relatively stable human traits, it was used in this study as a dependent variable to see if ATS tendencies were influenced in any way by a particular combination of our explicitly manipulated independent variables (IVs, appearance and agency) or by participant gender.

Procedure

The study was completed in 1 experimental session, following the sequence of the materials listed above: pretest, experimental manipulation, and posttest. In order, the pretest measures were the Desirability of Control Scale, the UCLA Loneliness Scale, the Need for Cognition Scale, and the Marlowe-Crowne Social Desirability scale. After the pretest measures, the experimenter read participants either the autonomous or nonautonomous story, depending on the agency manipulation condition to which the participant was assigned. Then, participants watched 1 of 2 videos, depending on their appearance manipulation condition, showing either a humanoid or a nonhumanoid robot interacting with a confederate actor. Finally, participants completed the posttest anthropomorphism measures, which, in order, were the Fussel adjective checklist, the PERNOD scale, the Epley et al anthropomorphic items, and finally the ATS. Upon completion of the third phase, the experimenter fully debriefed the participants before dismissal.

Data Reduction and Analyses

All scales and measures used in this study were scored for individual participants following the procedures described in the articles in which they were originally published. All dependent variables reported in this study were examined with the Shapiro-Wilk test to verify they were normally distributed within each of the 4 separate groups formed by the 2 manipulated IVs, appearance and agency. Accordingly, parametric tests were used for the analysis of the data collected in this study, as they are the most powerful means to assess the effects of the IVs [40]. For these analyses, both appearance and agency IVs were treated as between subject factors. Participant gender was also included in these analyses as a third, between-subjects factor to determine how the IVs affected both males and females in our study. Therefore, each analysis conformed to a 2 (appearance: humanoid vs nonhumanoid) × 2 (agency: autonomous vs nonautonomous) × 2 (participant gender: male vs female) analysis of variance (ANOVA) plan. Effects were considered significant in any ANOVA with P values ≤.05. Effect sizes were calculated in all ANOVAs as partial eta squared (η^2_p) to determine the degree of association between the variables. Partial eta squared values between 0.01 and 0.06 are considered small effects, between 0.06 and 0.14 are considered medium effects, and above 0.14 are large effects [34]. All significant interactions in the ANOVAs were followed up with individual group comparisons, and the Bonferroni procedure was applied to correct for multiple comparisons.

Results

Pretest Measure Analyses

Means, SDs, and group size for each of the separate groups in this study are provided in Table 2 for all 4 of the pretest measures. Separate 2 (appearance) × 2 (agency) × 2 (gender) preliminary analyses were conducted for each pretest measure to determine if there were any initial differences among groups on Desire for Control, Loneliness, Need for Cognition, or the Marlowe-Crowne scales. Results indicated no significant main effects or interactions for any of the pretest measures, with the exception of a main effect of gender within the Desire for Control Scale, $F_{1,91}$=5.62, P=.019, η^2_p=0.06, with males showing a greater desire for control than females, a finding consistent with the original work of Burger and Cooper [25].

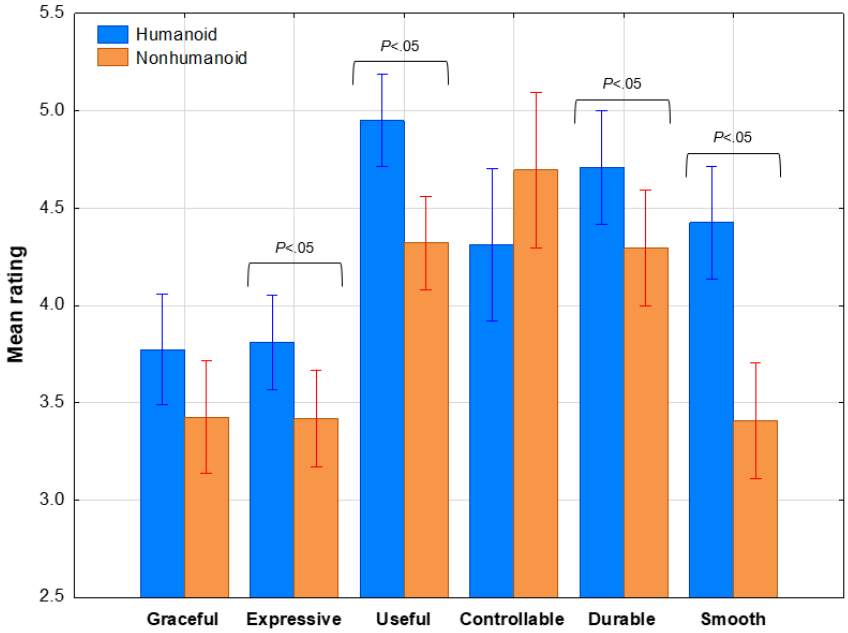
Table 2. Means, SDs, and group size for each of the groups in this study for all 4 of the pretest measures employed.

Appearance, agency, and gender subgroups			Desire for Control, mean (SD)	Loneliness, mean (SD)	Need for Cognition, mean (SD)	Marlowe-Crowne, mean (SD)
Appearance	Agency	Gender (N)				
Humanoid	Autonomous	Male (13)	105.7 (9.1)	38.3 (6.7)	113.1 (26.1)	4.8 (2.9)
Humanoid	Autonomous	Female (12)	101.0 (12.5)	39.8 (6.2)	116.1 (12.0)	4.5 (2.5)
Humanoid	Nonautonomous	Male (12)	103.3 (12.2)	37.8 (6.2)	106.3 (28.8)	5.2 (2.7)
Humanoid	Nonautonomous	Female (13)	97.5 (11.4)	39.2 (8.1)	101.8 (25.6)	5.8 (2.9)
Nonhumanoid	Autonomous	Male (13)	99.6 (13.8)	39.8 (6.7)	112.4 (19.9)	3.8 (1.7)
Nonhumanoid	Autonomous	Female (12)	96.7 (9.6)	39.1 (5.4)	110.4 (16.0)	4.1 (1.8)
Nonhumanoid	Nonautonomous	Male (14)	103.0 (10.3)	37.4 (9.5)	112.3 (21.1)	5.7 (3.4)
Nonhumanoid	Nonautonomous	Female (10)	95.3 (8.1)	39.7 (9.2)	102.5 (9.8)	3.4 (2.8)

Perception of the Humanoid and Nonhumanoid Robots

Figure 3 shows the mean rating on each of the 6 PERNOD subscales as a function of robot appearance (humanoid vs nonhumanoid). As is evident from this graph, the humanoid robot was perceived differently than the nonhumanoid on all dimensions. Generally, the humanoid morphology was associated with higher ratings on all subscales except *controllable*, where the opposite was true.

Figure 3. Mean rating on each perception to humanoid subscale as a function of robot appearance. Bars represent SE of the means.



The visual impressions in Figure 3 were confirmed by the results of an appearance × agency × gender multivariate analysis of variance (MANOVA) in which the 6 subscales were treated as multiple dependent measures. The main effect of appearance was significant in this analysis with a large effect size, $F_{6,86}$=6.59, P<.001, η^2_p=0.31. Separate follow up appearance × agency × gender ANOVAs were performed to see if the appearance effect was significant for each subscale. The results of these tests revealed that only appearance effects were significant for the *expressive* ($F_{1,91}$=5.9; P=.03), *useful* ($F_{1,91}$=13.75; P<.001), *durable* ($F_{1,91}$=3.83; P=.05), and *smooth* ($F_{1,91}$=23.78; P<.001) subscales, marginally significant for the *graceful* subscale (P=.09) but not significant for the *controllable* subscale. No other main effects or interactions were significant. These findings indicate that participants did perceive the humanoid and nonhumanoid robots differently, as expected, which confirms the effectiveness of our appearance manipulation.

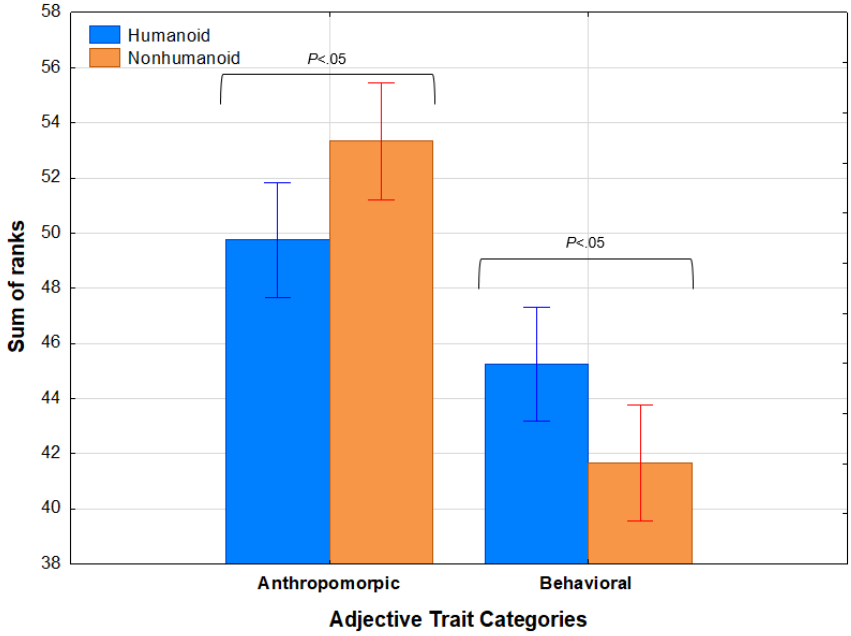
Measures of Anthropomorphism

The Epley et al Adjectives

Figure 4 depicts the reversed mean rank sums for each category of traits as a function of robot appearance. As is evident in Figure 4, generally higher rank sums for anthropomorphic than for behavioral traits were observed, meaning that participants believed that anthropomorphic traits were generally more applicable to both types of robots than were behavioral traits. However, within each trait category, appearance made a difference. Anthropomorphic traits were ranked higher for (were more applicable to) the nonhumanoid robot, whereas the behavioral traits were ranked higher for (were more applicable to) the humanoid robot.

Statistical confirmation for these observations was provided by an appearance × agency × gender × trait category ANOVA, which revealed a significant main effect of trait category, $F_{1,91}$=29.45, P<.001, η^2_p=0.24, as well as an appearance × category interaction, $F_{1,91}$=5.76, P=.02, η^2_p=0.06. Separate follow up appearance × agency × gender ANOVAs for each trait category revealed that the main effect of appearance was significant for both anthropomorphic $F_{1,91}$=5.78, P=.02 and behavioral $F_{1,91}$=5.76, P=.02 trait categories. For these analyses, neither gender nor agency were significant factors. These findings indicate that, contrary to our expectations, participants in this study were more willing to attribute anthropomorphic (ie, human like) traits to the nonhumanoid robot.

Figure 4. Mean rank sums for each category of Epley trait adjectives as a function of robot appearance. A higher sum of ranks indicates more applicable traits. Bars represent SE of the means.



The Fussel Adjective Checklist

Due to the multiple categories within the Fussel adjective checklist, statistical information is provided below for each category. These categories are as follows: human adjectives, robotic adjectives, and false filler and gender adjectives.

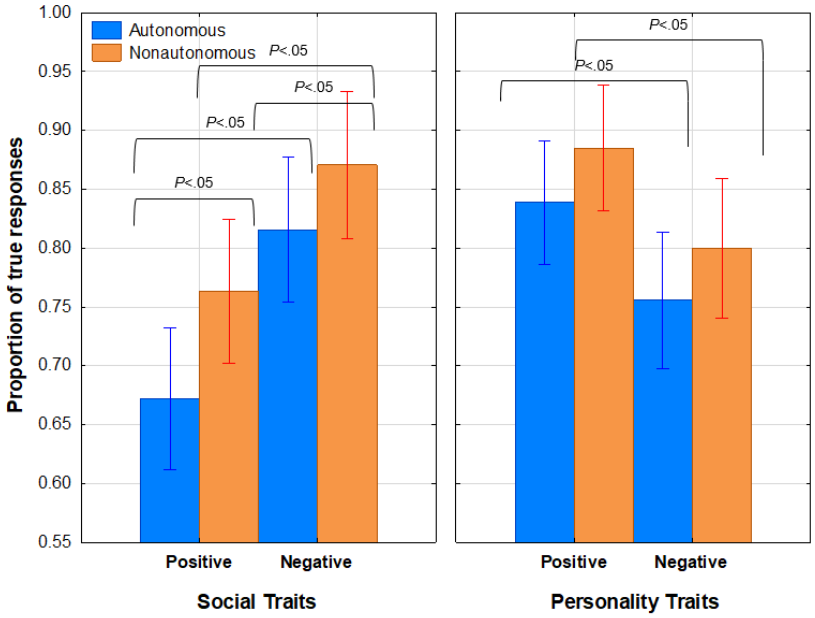
Human Adjectives

As shown in Table 1, the human adjectives were organized into 2 main categories (social and personality), each with a positive and negative subdivision. To examine the effects of our IVs on these 4 human adjective categories and subcategories, we applied an appearance × agency × participant gender × category (social vs personality) × valence (positive vs negative) ANOVA to the proportions of true adjectives in each category. For this analysis, category and valence were both within subject factors. This overall analysis revealed only a significant main effect for agency, $F_{1,91}$=4.88, P=.03, η^2_p=0.05, along with a significant category × valence interaction, $F_{1,91}$=50.57, P<.001, η^2_p=0.37.

Figure 5 shows the mean proportion of true responses as a function of agency, category, and valence. This graph illustrates that participants provided a higher proportion of true responses in each adjective category and subcategory for the nonautonomous robot. Moreover, it is clear that the category × valence interaction resulted from the reversal of the valence effect across categories. That is, for the 2 social adjective categories, participants ascribed more negative than positive attributes to the robot under both agency conditions, whereas the opposite was true for the 2 personality adjective categories.

To verify the basis for the interaction shown in Figure 5, appearance × agency × gender × valence ANOVAs were conducted separately for both social and personality adjective categories. For the social items analysis, the main effect of agency was significant, $F_{1,91}$=5.30, P=.02, η^2_p=0.06, as was the main effect of valence (positive vs negative), $F_{1,91}$=27.62, P<.001, η^2_p=0.23. For the personality items, only the main effect of valence was significant, F $1_{,91}$=7.35, P=.003, η^2_p=0.07. These outcomes verify that the valence effect was significant for both social and personality adjectives, but opposite in direction across categories. In addition, the effect of agency was arithmetically similar for both categories, but only reached significance for social adjectives. However, the fact that participants were more willing to attribute human like social qualities to the nonautonomous robot confirms that the agency manipulation was effective for this adjective category but also contradicts our original expectation that the autonomous robot condition would be perceived as the most human like.

Figure 5. Mean proportion of true responses for Fussel human adjective categories as a function of agency (autonomous or nonautonomous), category (social or personality), and valence (positive or negative). Bars represent SE of the means.

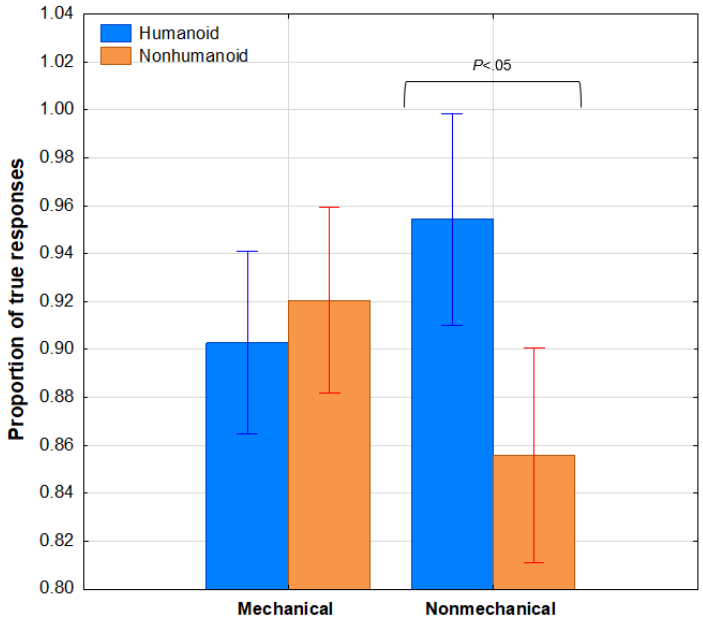


Robotic Adjectives

As shown in Table 1, the robotic adjectives from the original Fussel adjective checklist were subdivided into those that obviously referred to the mechanical characteristics of a robot and those that did not. Figure 6 shows the mean proportion of true responses to the mechanical and nonmechanical robotic adjectives as a function of robot appearance. What is apparent from this graph is that robot appearance made a difference only for the nonmechanical characteristics of the robotic adjectives.

The visual impressions evident in Figure 6 were confirmed by the results of an appearance × agency × participant gender × robotic adjective category (mechanical vs nonmechanical) ANOVA in which the appearance × robotic adjective category interaction was significant, $F_{1,91}$=10.58, P=.001, η^2_p=0.10. Follow up tests showed that the interaction was because of a significant difference between appearance conditions only for the nonmechanical adjective category (P=.003). These results from the analysis of the robotic adjective category indicate that both robot morphologies were perceived to be equally mechanical, but the humanoid robot was perceived to be different from the nonhumanoid in nonmechanical ways (ie, portability).

Figure 6. Mean proportion of true responses to the Fussel adjective checklist mechanical and nonmechanical robotic adjectives as a function of robot appearance. Bars represent SE of the means.



False Filler and Gender Adjectives

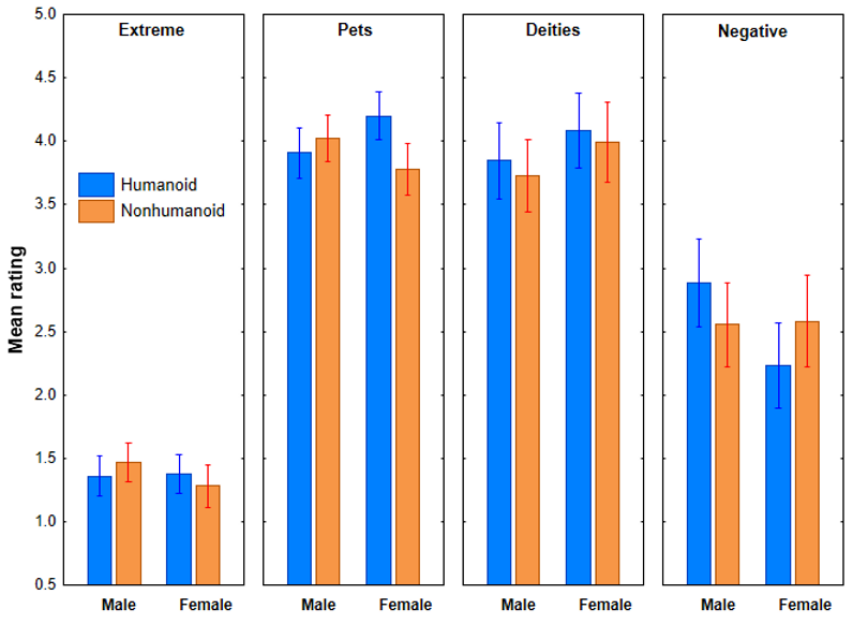
The 2 remaining adjective categories in Table 1 were also examined. Participants in all conditions reported a low percentage of true responses for adjectives in the false filler category, which was expected based on the fact that these adjectives were selected specifically because they did not apply to humans or robots [38]. An appearance × agency × gender ANOVA of these items failed to reveal any significant effects. For the gender adjective category, a greater proportion of participants perceived the humanoid robot as male compared with the nonhumanoid robot, whereas the tendency to perceive either robot as female was equivalent for the 2 appearance categories. An analysis of the gender adjective proportions using an appearance × agency × participant gender × robot gender ANOVA revealed only a significant appearance × robot gender interaction, $F_{1,91}$=5.83, P=.02, $\eta 2_p$=0.06. Follow up tests showed that the interaction was because of a significant difference between appearance conditions for the male robot gender category (P=.03), but not for the female category.

Anthropomorphic Tendencies Scale

Figure 7 shows mean ratings for each ATS subscale as a function of robot appearance and participant gender. A total of 3 observations are apparent from this graph. First, participants generally agreed with statements reflecting anthropomorphism of pets and deities but disagreed with statements of extreme anthropomorphism. However, participants were more neutral about negative anthropomorphism statement. Second, ignoring robot appearance, male and female participants reported about the same levels of agreement with extreme anthropomorphism and anthropomorphism of pets but differed slightly in agreement with statements of anthropomorphism of deities and negative anthropomorphism. Third, the effect of robot appearance was different for males and females, most notably for anthropomorphism of pets and negative anthropomorphism.

To examine the trends in Figure 7, an appearance × agency × gender × ATS subscale (extreme vs pets vs deities vs negative) MANOVA was conducted, where the subscale scores were treated as separate dependent variables. This analysis revealed only a significant appearance × gender interaction, $F_{4,86}$=4.10, P=.004, η^2_p=0.16. To better understand the appearance × gender interaction in the overall MANOVA, separate appearance × agency × gender interactions were conducted for each ATS subscale. The appearance × gender interactions were significant in these analyses only for anthropomorphism of pets, $F_{1,89}$=7.47, P=.007, η^2_p=0.07 and negative anthropomorphism, $F_{1,89}$=3.89, P=.05, η^2_p=0.04. Moreover, the main effect of gender was marginally significant for anthropomorphism of deities, $F_{1,89}$=2.79, P=.09, η^2_p=0.03. Follow up individual group comparisons revealed that females differed significantly in their reported anthropomorphism of pets as a function of robot appearance (P=.02), and males expressed significantly more negative anthropomorphism than females under the humanoid appearance condition (P<.001). These results indicate, once again, that robot appearance was an effective variable in this study, at least for female versus male expressions of anthropomorphism of pets and negative anthropomorphism.

Figure 7. Mean ratings for each Anthropomorphic Tendencies Scale subscale as a function of robot appearance and participant gender. Bars represent SE of the means.



Discussion

We can summarize the results of this study in the context of our original study purposes and hypotheses.

Hypothesis 1: The Effect of Humanoid Appearance on Robot Anthropomorphism

In accord with the SEEK model of anthropomorphism [1], we hypothesized that a robot having a more human like morphology would provoke greater availability and use of homocentric knowledge than a less human like robot, which in turn might

lead participants to engage in more anthropomorphism toward the humanoid than the nonhumanoid robot. To test this possibility, we employed an *appearance* manipulation involving robots with either a humanoid or a nonhumanoid form. Participants were asked to observe a videotaped interaction between 1 of these robots and a human and then complete several different measures of anthropomorphism that have been used in previous studies.

Results with the PERNOD scale showed that participants perceived the humanoid robot to be significantly more useful, expressive, graceful, and durable than the nonhumanoid robot and marginally more pleasing in appearance. These results demonstrated that the appearance manipulation made a difference on this measure and may also support the first hypothesis to the extent that the PERNOD expressive subscale (ie, the extent to which the robot communicates emotion and/or friendliness) is an indicator of anthropomorphic attributions. Of course, these findings do not tell us specifically which aspects of appearance were responsible for the differences revealed by this measure. Additional research will be needed to expand on these findings.

However, findings from the Epley and Fussel adjective measures of anthropomorphism [2,38] do not seem to support hypothesis 1. Although they revealed significant effects of the robot appearance factor, those effects were opposite in direction to our expectations. In both cases, participants were more willing to attribute human like traits to the nonhumanoid robot than to its humanoid counterpart.

It is not clear how to interpret our findings with the ATS in relation to hypothesis 1. These results were complicated by an appearance × gender interaction in which only females reported more anthropomorphism of pets under the humanoid appearance condition. In addition, females reported less negative anthropomorphism than males under the humanoid condition. At best, the ATS offers only limited support for our original hypothesis that the humanoid robot should provoke more anthropomorphic tendencies in our participants as the significant effects of appearance we observed with this measure were only for females.

Hypothesis 2: The Effect of Agency on Robot Anthropomorphism

Our second hypothesis was that participants would anthropomorphize an autonomous robot more than one that was not autonomous. This expectation, derived from the SEEK model, was based on the idea that autonomy would allow participants to better understand and explain the robot's behavior by applying their own anthropocentric knowledge to it. To test this possibility, we employed an *agency* manipulation in which participants were told that the robot they were about to see in the video either was sophisticated and capable of acting on its own or was being controlled by someone in another room. In contrast to a straightforward explanation of the nonautonomous robot's behavior as being remotely controlled, we expected that participants would be more inclined to interpret the autonomous robot's actions in more human like terms (eg, being friendly or sociable).

Interestingly, however, the only significant effects of agency obtained in this study ran counter to our expectations. For the Fussel adjective checklist items [38], we found that participants were more likely to make attributions of both positive (eg, friendly and sociable) and negative (eg, rude and aggressive) human sociality traits to the nonautonomous rather than the autonomous robot, regardless of appearance. The same trend appeared in the attributions of personality traits (eg, organized and distractible) but these findings did not achieve statistical significance. Therefore, once again as with appearance, we are left to wonder why our agency manipulation did not work as expected.

The manipulations in this study of appearance and agency both seemed to have independent influences upon how participants perceived the robot to which they were exposed. However, contrary to expectations based on the elicited agent knowledge factor in the SEEK model of anthropomorphism, participants were less willing to make attributions of human like social or personality traits toward a robot when its morphology had more human like features but were more willing to make those attributions when they were told that the robot was being remotely controlled by a person rather than acting on its own. As these influences did not appear to interact, it is appropriate to consider separate explanations for these somewhat surprising effects of both our IVs.

The Overall Size Factor

One possible explanation of the unexpected effects of robot appearance in our study is based on the overall size differential between the two robots. As the Nao humanoid robot was physically smaller in stature than the PeopleBot nonhumanoid platform, it is possible that the propensity to make human like attributions could have been influenced by this factor. We took two precautions to mitigate the possible effects of size difference in this study. The first was to expose participants to only one of our robots using prerecorded videos rather than using actual physical exposure to the robots. On the basis of the findings of an earlier study [32], we expected video exposure to yield comparable effects with actual physical exposure, and, although not eliminating the perception of size, video exposure might also mitigate the perception of apparent size relative to actual exposure, especially when participants do not see a second robot to which they can compare the first. A second precaution, noted earlier in the appearance manipulation video section, was that we positioned both robots for filming so that their heads were approximately equidistant from the floor and approximately at the same viewing angle with respect to the confederate used in the video.

Despite these precautions, however, there is at least modest evidence that robot size registered with our participants. For example, the right-hand portion of Figure 5 reveals that the humanoid robot differed significantly from its nonhumanoid counterpart in terms of nonmechanical attributes such as portability, a finding that may be partly a reflection of overall size. In addition, there were significant differences shown in Figure 3 in favor of the humanoid robot being perceived as the more useful platform, which also might be at least partly size related. Perhaps these differences are a reflection of the fact

that participants considered the humanoid Nao to be more toy like (as it was perceived as more portable and useful) than the nonhumanoid PeopleBot. Possibly, these impressions influenced anthropomorphism tendencies in our participants. Further work needs to build on previous anthropomorphism research [33,35,36,41-43] not only in exploring the possible effects of robot height but also examining overall size. Undoubtedly, these considerations will prove to be quite relevant to a fuller understanding of HRI.

Indirect Agency

A different possible explanation to account for our counterintuitive finding that a remote-controlled robot was perceived as more human like than an *autonomous* robot is that participants in the nonautonomous conditions in this study actually made attributions toward what might be called the *indirect agent*. As participants in this condition were told that a person was controlling the robot, it is very possible that they viewed the nonautonomous robot merely as a kind of interface for a remote controlling human agent. Thus, attributions of humanness directed at the robot really might have been intended for the person thought to be behind the machine.

Other work indirectly supports the notion of indirect agency by showing that children are more empathetic toward teleoperated robots [44], people feel more social presence with teleoperated robots [45], and people have identified a teleoperated search and rescue robot as being warmer, safer, and more attentive than an autonomous robot [46]. In this study, participants were more willing to make human like attributions of positive sociality (ie, friendly, sensitive, and caring) to the nonautonomous robot being remotely controlled by a human than to an autonomous robot supposedly acting on its own. Quite possibly, this means that participants were making these human-like attributions toward the operator behind the robot rather than toward the robot itself.

The use of remotely controlled or teleoperated robots is a common strategy in studies of HRI that has come to be known as the *Wizard of Oz* paradigm [6,47,48]. The possibility that, under these circumstances, the robot might be viewed by participants merely as a surrogate for the human behind the scenes is a feature of this paradigm that has not received much attention, largely because the existence of the *wizard* usually is hidden from participants. Nonetheless, it is clear that we need to have a better understanding of when and how participants look past the machines with which they interact to the people they think are controlling those machines, or maybe even to those they think created or programmed them.

Gender Effects

A secondary purpose of this study was to examine how males and females reacted to the manipulations in this study and responded to the various scales employed to measure robot perception and anthropomorphism. However, very few participant gender differences were observed. This finding suggests that male and female participants in this study perceived and made attributions about the 2 robotic platforms in essentially the same way. However, there was limited evidence that females in the humanoid robot condition reported more anthropomorphism toward pets and less negative anthropomorphism than females in the nonhumanoid robot condition.

Conclusions

This study clearly indicated that physical robot appearance makes a difference in how people perceive robot platforms. The Nao robot in this study was perceived as more useful, expressive, graceful, and durable, and possibly smoother than the PeopleBot. These perceptions are important to document and explore in relation to how humans interact with different robotic platforms as well as what preferences humans might exhibit for interacting with 1 platform over another. Moreover, the finding in this study that the Nao humanoid robot was perceived as more masculine than the PeopleBot also may prove important in situations where perceived robot gender can influence HRI. These findings also suggest that participants made indirect agency attributions to the humans operating behind the robot, a finding of potential widespread significance in HRI. The robustness and boundary conditions for such indirect attributions need to be further explored and better understood.

Finally, we wish to note that general theories of anthropomorphism, such as the SEEK model [1], need to be more fully explored and tested in the context of HRI. In this study, we tested only 1 factor in SEEK model, elicited agent knowledge, and obtained some unexpected findings. It is important to understand how anthropomorphizing robots may be similar to or different from the anthropomorphism of other nonhuman entities. This work, as well as that of other SEEK model studies [5,12,23,24], represents important initial steps toward this goal.

Acknowledgments

The authors would like to thank Dr Laurel Riek for providing helpful comments on an earlier draft of this paper. They also wish to thank the undergraduate research assistants, past and present, of the *e* Motion & *e* Cognition laboratory for their help in collecting the present data.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Both agency priming scripts.

[DOCX File, 11KB - humanfactors_v6i2e12629_app1.docx]

Multimedia Appendix 2

Video of humanoid robot and confederate interaction.

[MP4 File (MP4 Video), 174MB - humanfactors_v6i2e12629_app2.mp4]

Multimedia Appendix 3

Video of nonhumanoid robot and confederate interaction.

[MP4 File (MP4 Video), 174MB - humanfactors_v6i2e12629_app3.mp4]

References

1. Epley N, Waytz A, Cacioppo J. On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 2007 Oct;114(4):864-886. [doi: 10.1037/0033-295X.114.4.864] [Medline: 17907867]
2. Epley N, Waytz A, Akalis S, Cacioppo J. When we need a human: motivational determinants of anthropomorphism. Social Cognition 2008 Apr;26(2):143-155. [doi: 10.1521/soco.2008.26.2.143]
3. Nass C, Lee KM. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. J Exp Psychol Appl 2001;7(3):171-181. [doi: 10.1037//1076-898X.7.3.171]
4. Duffy B. Anthropomorphism and the social robot. Rob Auton Syst 2003;42(3-4):177-190. [doi: 10.1016/S0921-8890(02)00374-3]
5. Eyssel F, Kuchenbrandt D, Bobinger S. Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. : ACM Press; 2011 Presented at: 6th International Conference on Human Robot Interaction; 2011; Lausanne, Switzerland p. 61-68. [doi: 10.1145/1957656.1957673]
6. Villano M, Crowell C, Wier K, Tang K, Thomas B, Shea N, et al. DOMER: a wizard of oz interface for using interactive robots to scaffold social skills for children with autism spectrum disorders. : IEEE Press; 2011 Presented at: 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2011; Lausanne, Switzerland p. 279-280. [doi: 10.1145/1957656.1957770]
7. Powers A, Kiesler S. The advisor robot: tracing people's mental model from a robot's physical attributes. : ACM Press; 2006 Presented at: ACM SIGCHI/SIGART conference on Human-robot interaction; 2006; Salt Lake City, Utah p. 218-225. [doi: 10.1145/1121241.1121280]
8. Riek L, Rabinowitch T, Chakrabarti B, Robinson P. How anthropomorphism affects empathy toward robots. : ACM Press; 2009 Presented at: 4th ACM/IEEE international conference on Human robot interaction; 2009; La Jolla, California p. 43-48. [doi: 10.1145/1514095.1514158]
9. Hinds P, Roberts T, Jones H. Whose job is it anyway? A study of human-robot interaction in a collaborative task. Int J Hum Comput Interact 2004;19(1):151-181. [doi: 10.1207/s15327051hci1901&2_7]
10. Goetz J, Kiesler S, Powers A. Matching robot appearance and behavior to tasks to improve human-robot cooperation. : IEEE; 2003 Presented at: The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003; 2003; Millbrae, California p. 55-60. [doi: 10.1109/ROMAN.2003.1251796]
11. Paepcke S, Takayama L. Judging a bot by its cover: An experiment on expectation setting for personal robots. : IEEE Press; 2010 Presented at: 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2010; Osaka, Japan p. 45-52. [doi: 10.1145/1734454.1734472]
12. Eyssel F, Kuchenbrandt D, Hegel F, de Ruiter L. Activating elicited agent knowledge: How robot and user features shape the perception of social robots. : IEEE Press; 2012 Presented at: The 21st IEEE International Symposium on Robot and Human Interactive Communication; 2012; Paris, France p. 851-857. [doi: 10.1109/ROMAN.2012.6343858]
13. Crowell C, Scheutz M, Schermerhorn P, Villano M. Gendered voice and robot entities: perceptions and reactions of male and female subjects. : IEEE Press; 2009 Presented at: 2009 IEEE/RSJ international conference on Intelligent robots and systems; 2009; St. Louis, Missouri p. 3735-3741. [doi: 10.1109/IROS.2009.5354204]
14. Syrdal D, Koay K, Walters M, Dautenhahn K. A personalized robot companion? - The role of individual differences on spatial preferences in HRI scenarios. : IEEE Press; 2008 Presented at: The 16th IEEE International Symposium on Robot and Human Interactive Communication; 2007; Jeju, South Korea p. 1143-1148. [doi: 10.1109/ROMAN.2007.4415252]
15. Nomura T, Suzuki T, Kanda T, Kato K. Measurement of negative attitudes toward robots. Interact Stud 2006 Nov 15;7(3):437-454. [doi: 10.1075/is.7.3.14nom]
16. Nomura T, Kanda T, Suzuki T, Kato K. Exploratory investigation into influence of negative attitudes toward robots on human-robot interaction. In: Mobile robots: towards new applications. London: IntechOpen; 2006.

17. Nomura T, Kanda T, Suzuki T. Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. AI Soc 2005 Aug 26;20(2):138-150. [doi: 10.1007/s00146-005-0012-7]
18. Nomura T, Suzuki T, Kanda T, Kato K. American Association for Artificial Intelligence. 2006. Altered attitudes of people towards robots: investigation through the negative attitudes towards robots scale URL: http://www.aaai.org/Papers/Workshops/2006/WS-06-09/WS06-09-006.pdf [accessed 2019-04-16] [WebCite Cache ID 77gD6F5nr]
19. Halpern D, Katz J. Unveiling robotophobia and cyber-dystopianism: the role of gender, technology and religion on attitudes towards robots. : IEEE Press; 2012 Presented at: The 7th annual ACM/IEEE international conference on Human-Robot Interaction; 2012; Boston, Massachusetts p. 139-140. [doi: 10.1145/2157689.2157724]
20. May DC, Holler KJ, Bethel CL, Strawderman L, Carruth DW, Usher JM. Survey of factors for the prediction of comfort with a non-anthropomorphic robot in public spaces. Int J of Soc Robotics 2017 Jan 13;9(2):165-180. [doi: 10.1007/s12369-016-0390-7]
21. Kuo I, Rabindran J, Broadbent E, Lee Y, Kerse N, Stafford R, et al. Age and gender factors in user acceptance of healthcare robots. : IEEE; 2009 Presented at: The 18th IEEE International Symposium on Robot and Human Interactive Communication; 2009; Toyama, Japan p. 214-219. [doi: 10.1109/ROMAN.2009.5326292]
22. Chin M, Sims V, Clark B, Lopez G. Measuring individual differences in anthropomorphism toward machines and animals. Proc hum factors ergon soc annu meet 2004 Sep 1;48(11):1252-1255. [doi: 10.1177/154193120404801110]
23. Eyssel F, Kuchenbrandt D. Manipulating anthropomorphic inferences about NAO: the role of situational and dispositional aspects of effectance motivation. : IEEE; 2011 Presented at: 2011 RO-MAN; 2011; Atlanta, Georgia p. 467-472. [doi: 10.1109/ROMAN.2011.6005233]
24. Eyssel F, Reich N. Loneliness makes the heart grow fonder (of robots) — On the effects of loneliness on psychological anthropomorphism. 2013 Presented at: 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)s; 2013; Tokyo, Japan p. 121-122. [doi: 10.1109/HRI.2013.6483531]
25. Burger J, Cooper H. The desirability of control. Motiv Emot 1979 Dec;3(4):381-393. [doi: 10.1007/BF00994052]
26. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
27. Cacioppo JT, Petty RE, Kao CF. The efficient assessment of need for cognition. J Pers Assess 1984 Jun;48(3):306-307. [doi: 10.1207/s15327752jpa4803_13] [Medline: 16367530]
28. Crowne D, Marlowe D. A new scale of social desirability independent of psychopathology. J Consult Psychol 1960;24(4):349-354. [doi: 10.1037/h0047358]
29. Reynolds W. Development of reliable and valid short forms of the marlowe-crowne social desirability scale. J Clin Psychol 1982 Jan;38(1):119-125. [doi: 10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I]
30. Negi S, Arai T, Inoue K, Ujiie Y, Takubo T. Psychological assessment of humanoid robot appearance and performance using virtual reality. : IEEE Press; 2008 Presented at: The 17th IEEE International Symposium on Robot and Human Interactive Communication; 2008; Munich, Germany. [doi: 10.1109/ROMAN.2008.4600752]
31. Kamide H, Yasumoto M, Mae Y, Takubo T, Ohara K, Arai T. Comparative evaluation of virtual and real humanoid with robot-oriented psychology scale. : IEEE Press; 2011 Presented at: International Conference on Robotics and Automation; 2011; Shanghai, China p. 599-604. [doi: 10.1109/ICRA.2011.5979893]
32. Woods S, Walters M, Koay KL, Dautenhahn K. Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials. : IEEE; 2007 Presented at: The 15th IEEE International Symposium on Robot and Human Interactive Communication; 2006; Hatfiel, United Kingdom p. 219-232. [doi: 10.1109/ROMAN.2006.314394]
33. Jouppi N, Thomas S. Telepresence systems with automatic preservation of user head height, local rotation, and remote translation. 2006 Presented at: IEEE International Conference on Robotics and Automation; 2005; Barcelona, Spain p. 62-68. [doi: 10.1109/ROBOT.2005.1570097]
34. Cohen J. Statistical Power Analysis For The Behavioral Sciences (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
35. Walters M, Dautenhahn K, te Boekhorst BR, Koay K, Syrdal D, Nehaniv C. An empirical framework for human-robot proxemics. 2009 Presented at: Proc New Front Hum Robot Inter; 2009; Edinburgh, Scotland p. 144-149 URL: http://hdl.handle.net/2299/9670
36. Walters M. University of Hertfordshire.: University of Hertfordshire, Hertfordshire, United Kingdom; 2008. The design space for robot appearance behavior for social robot companions URL: https://uhra.herts.ac.uk/handle/2299/1806 [accessed 2019-04-16] [WebCite Cache ID 77gGVywfd]
37. Kamide H, Mae Y, Takubo T, Ohara K, Arai T. Development of a scale of perception to humanoid robots: PERNOD. : IEEE; 2010 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems; 2010; Taipei, Taiwan p. 5830-5835. [doi: 10.1109/IROS.2010.5648955]
38. Fussell S, Kiesler S, Setlock L, Yew V. How people anthropomorphize robots. : ACM Press; 2008 Presented at: 3rd ACM/IEEE international conference on Human robot interaction; 2008; Amsterdam, The Netherlands p. 145-152. [doi: 10.1145/1349822.1349842]

39. Chin MG, Yordon RE, Clark BR, Ballion T, Dolezal MJ, Shumaker R, et al. Developing and Anthropomorphic Tendencies Scale. 2016 Nov 05 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2005; Washington, District of Columbia p. 1266-1268. [doi: 10.1177/154193120504901311]
40. Maxwell S, Delaney H, Kelley K. Designing experiments and analyzing data: a model comparison perspective (2nd Ed). New York, New York: Wadsworth Publishing Company; 2004.
41. Rae I, Takayama L, Mutlu B. The influence of height in robot-mediated communication. 2013 Presented at: 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2013; Tokyo, Japan p. 1-8. [doi: 10.1109/HRI.2013.6483495]
42. Lee M, Forlizzi J, Rybski P, Crabbe F, Chung W, Finkle J, et al. The snackbot: documenting the design of a robot for long-term human-robot interaction. : ACM; 2009 Presented at: 4th ACM/IEEE international conference on Human robot interaction; 2009; La Jolla, California p. 7-14. [doi: 10.1145/1514095.1514100]
43. Koay K, Syrdal D, Walters M, Dautenhahn K. Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study. : IEEE Press; 2008 Presented at: The 16th IEEE International Symposium on Robot and Human Interactive Communication; 2007; Jeju, South Korea p. 564-569. [doi: 10.1109/ROMAN.2007.4415149]
44. Kwak S, Kim Y, Kim E, Shin C, Cho K. What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. : IEEE Press; 2013 Presented at: 2013 IEEE RO-MAN; 2013; Gyeongju, South Korea p. 180-185. [doi: 10.1109/ROMAN.2013.6628441]
45. Choi J, Kim Y, Kwak S. The autonomy levels and the human intervention levels of robots: The impact of robot types in human-robot interaction. 2014 Presented at: The 23rd IEEE International Symposium on Robot and Human Interactive Communication; 2014; Edinburgh, United Kingdom p. 1069-1074. [doi: 10.1109/ROMAN.2014.6926394]
46. Dole L, Sirkin D, Currano R, Murphy R, Nass CI. Where to look and who to be designing attention and identity for search-and-rescue robots. : IEEE Press; 2013 Presented at: 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2013; Tokyo, Japan p. 119-120. [doi: 10.1109/HRI.2013.6483530]
47. Green A, Hüttenrauch H, Eklundh K. Applying the wizard-of-oz framework to cooperative service discovery and configuration. : IEEE Press; 2005 Presented at: 3th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759); 2004; Kurashiki, Okayama, Japan p. 20-22. [doi: 10.1109/ROMAN.2004.1374824]
48. Riek L. Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. J Hum Robot Interact 2012 Aug 01;1(1):119-136. [doi: 10.5898/JHRI.1.1.Riek]

Abbreviations

ANOVA: analysis of variance
ATS: Anthropomorphic Tendencies Scale
HRI: human robot interaction
IVs: independent variables
MANOVA: multivariate analysis of variance
PERNOD: Perception of Humanoid Robots Scale
SEEK: Sociality, Effectance, Elicited agent Knowledge

Edited by C Eichenberg; submitted 31.10.18; peer-reviewed by A Cesta, S Rossi; comments to author 14.01.19; revised version received 05.03.19; accepted 06.04.19; published 10.05.19.

Please cite as:
Crowell CR, Deska JC, Villano M, Zenk J, Roddy Jr JT
Anthropomorphism of Robots: Study of Appearance and Agency
JMIR Hum Factors 2019;6(2):e12629
URL: http://humanfactors.jmir.org/2019/2/e12629/
doi:10.2196/12629
PMID:31094323

Original Paper

The Effect of Age on Electronic Health Literacy: Mixed-Method Study

Wan-Chen Hsu[1], BSc, MA, PhD

Center for Teaching and Learning Development, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

Corresponding Author:
Wan-Chen Hsu, BSc, MA, PhD
Center for Teaching and Learning Development
National Kaohsiung University of Science and Technology
No 415, Jiangong Rd
Sanmin Dist
Kaohsiung, 80778
Taiwan
Phone: 886 73814526 ext 31171
Email: wanchen@nkust.edu.tw

Abstract

Background: The world's internet penetration rate is increasing yearly; approximately 25% of the world's population are internet users. In Asia, Taiwan has the fifth highest internet usage, and has an internet penetration rate higher than the world average. Electronic health (eHealth) literacy is the ability to read, understand, and utilize Web health information. eHealth literacy is gaining attention worldwide.

Objective: This study aimed compare the differences in eHealth literacy between traditional college students (aged between 18 and 22 years) and older adult students (aged between 55 and 72 years). It also summarizes the experiences and performances of these 2 groups in terms of searching online health-related information.

Methods: A mixed-method approach was used, including questionnaire surveys and interviews. A total of 208 respondents were interviewed: 65 traditional college students (31.3%) and 143 older adult students (68.7%). The results of the interviews were used to compare the eHealth literacy scores of the 2 groups.

Results: There were significant differences in the overall eHealth literacy scores (t_{207}=2.98; P=.001) and the functional eHealth literacy dimension (t_{207}=12.17; P<.001). The findings showed a significant gap in eHealth literacy between the 2 groups. Most participants believed that online health information could be largely read and understood. However, they were skeptical about the quality of the information and noted that it consisted of either subjective judgments or objective standards.

Conclusions: Traditional college students preferred esthetically pleasing health information, whereas older adult students focused on its promotion. Furthermore, the first group often used websites for solving health problems, whereas the second group forwarded health information through communication software.

(JMIR Hum Factors 2019;6(2):e11480) doi:10.2196/11480

KEYWORDS

eHealth literacy; intergenerational relations; traditional college students; older adult students; mixed method

Introduction

Increased Internet Usage and the Issue of Electronic Health Literacy

Electronic health (eHealth) literacy is gaining increasing attention worldwide. Individuals with eHealth literacy have better health capital and can further promote the overall health and competitiveness of their countries. In 2017, the Taiwan Broadband Internet Usage Survey reported that results from 3153 valid sample analyses showed that individuals were accustomed to having an internet access rate of over 83%. Of these individuals, 84.7% agreed that "the use of the Internet has improved the quality of your life." However, 55.9% disagreed with the statement that "the use of the Internet can increase your trust in information" [1]. Thus, even if individuals possessed little knowledge of eHealth information, it was still possible for them to take appropriate action.

Different Age Groups Exhibit Unique Electronic Health Literacy Performances

Given the popularity of the internet, research has shifted focus toward the relevance of health literacy through the internet; as a result, "eHealth Literacy," as a field of research, has gradually received increased attention [2]. Eng [3] argues that *eHealth literacy* refers to the use of the internet to access information to improve or promote health, whereas Norman and Skinner [4] further suggest that *eHealth literacy* refers to the use of the internet to seek, understand, and evaluate health information and to use this information to address health problems.

Health literacy includes 3 dimensions [5], and it is the main constituent of eHealth literacy [2]. Chiang et al [6], using the definition of health literacy, divided the 3 dimensions into functional eHealth literacy, interactive eHealth literacy, and critical eHealth literacy. The first layer is *functionality*, which refers to the essential ability to read health information. The second layer is *interaction*, which refers to more advanced knowledge of the choices available in health information, including knowledge required to understand, integrate, and use that information and knowledge of a supportive, interactive environment that provides health information and other skills. The third layer is *criticality*, which is a more in-depth analysis of health information and involves both criticisms of the information and its application to health and the response to that criticism, resulting in better control over living conditions.

Related Research on Electronic Health Literacy

eHealth literacy has the potential to positively support consumers' health empowerment [7]. The Integrative Model of eHealth Use claims that macro-level disparities in social structure are connected to health disparities that arise because of micro-level factors such as eHealth literacy, motivation, and ability [2]. Few studies have explored the associations among individual factors such as gender, age, and college major in relation to eHealth literacy [8,9], as cross-group comparisons have yet to be investigated empirically. In Taiwan, which currently has the fifth-highest internet usage in Asia and has an internet penetration rate of 65.90% [10], college students constitute 1 of the groups that access internet health information more frequently. The proportion of older people using information and communication technology appears to be lower than that of other age groups, making it difficult to breach this digital wall. Currently, there are nearly 300 active aging learning centers in Taiwan, which are the leading institutions of learning for retired older adults. They are designed to receive students over the age of 55 years [11]. This leads to the question examined in this study: Does eHealth literacy overcome the generation gap? Specifically, is there a difference between eHealth literacy among traditional college students and older adult students? To answer these questions, this study examined the differences in eHealth literacy among older adult students (people over the age of 55 years) and among traditional college students (aged from 18 to 22 years, thus within the conventional age range for university undergraduates in Taiwan) to gain an in-depth understanding of the differences that exist across age groups.

Eysenbach and Köhler [12] explored the use of health information by internet users and found that participants' assessment of the quality of online health information included the authority, appearance, and layout of the source; advertising; readability; the presence (or absence) of links to other websites; website holder photos; contact boxes; website certification; content updates; quality badges; or other professional group support. There is limited literature on the experience of college students using online health information. Within that literature, studies have shown that traditional college students possess functional and interactive health literacy levels and seem to underperform at a higher level of critical health literacy. Relevant studies have shown that college students are confident that they can find, read, and understand online health information [8,13]. However, a high proportion of these students are less assured in their ability to discriminate between high- and low-quality health resources [13].

Compared with younger adults, older adults had less confidence in eHealth resources, information-seeking skills, and the ability to evaluate and act upon online health information [14]. Lee et al [15] note that increased age is a factor that is frequently associated with decreased levels of eHealth literacy. Older adults with chronic health conditions and those with lower levels of eHealth literacy were prone to unmet navigational needs, experiencing difficulties in finding online health information, and being less assured in their searching abilities [15]. An investigation of internet skills also found that older adults sometimes experienced problems when completing tasks that called upon operational and formal internet skills [16]. This included difficulties in understanding orientation within a website and identifying and using the browser address bar. A survey that addressed the health information–seeking behaviors of baby boomers and older adults found that an increase in age corresponded to a decrease in eHealth literacy scores [17]. However, in contrast to other research studies, Lee et al [15] found that the respondents were mostly confident regarding their ability to find and use internet-based health resources, although they were less confident in their ability to differentiate between high- and low-quality resources. The more relevant generational differences were based on quantitative research. The novelty of this study is that it takes into account both quantitative and qualitative research methods in coming to an understanding of the prevalence of eHealth literacy among participants of different generations.

In Taiwan, nearly half of traditional college students use internet health information as a conduit for self-diagnosis [18]. Moreover, in April 2018, the proportion of people over the age of 65 years reached 14% in Taiwan's population structure. Therefore, Taiwan has officially become an "aged society" [19] that pays attention to the current situation of traditional college students and elders in eHealth literacy and is more concerned with the overall competitiveness of the country in the future.

Taiwan has made improving the health literacy of adults a cornerstone of national health policy. At present, the research on health literacy–related topics mostly focuses on the preparation of measurement tools [9,20,21] and the current lack of intergenerational differences in adult eHealth literacy. This study used a mixed-method perspective to explore the eHealth

literacy of different groups of traditional college students and older adult students. Complementing the understanding of the current situation on eHealth literacy is the importance of exploring health literacy as a topic of national policy and the bridging of gaps in the research literature on the subject. Information literacy and health literacy are important connotations of eHealth literacy [4]. In study 1, *a quantitative analysis is performed to explore eHealth literacy between different age groups*. eHealth literacy has been investigated by the electronic Health Literacy Scale (eHLS) and clearly represented in participants' eHLS scores. These scores correspond to self-reported ability to find and evaluate online health information. Thus, study 2 focuses on participants' information-seeking behaviors, patterns, and preferences and on their skills in detecting online health information. To examine this issue further, this study also summarized the online health information–seeking behaviors, patterns, and preferences of both traditional college students and older adult students concerning skills, experiences, and performances. The purpose of this study was to compare the eHealth literacy scores of both groups of students. Moreover, by collecting the online health information of these groups through interviews and then summarizing this information, this study sought to read and understand what constitutes relevant experience for each group.

Methods

Overview

This study uses a sequential mixed-method design. Such a design uses qualitative and quantitative research sequentially, depending on the purpose or problem of the study. The purpose is both to attain "complementarity," such as the rationality of quantitative data in additional sampling, and to further research through qualitative study [22]. Accordingly, this study was carried out in 2 phases. The first phase, comprising the *quantitative* part of the study, was mainly used to screen respondents. The second phase consisted of an interview to collect the respondents' eHealth literacy data, which were analyzed *qualitatively*.

Recruitment

In this mixed-method study, the research process was divided into 2 phases. First, an eHealth literacy instrument (a questionnaire) was used to investigate the participants' current situation. In this first stage, 2 classes of traditional college students in the general education program were assessed. Moreover, 3 classes of older adult students (aged 55 years or above) who were enrolled in a university-affiliated, formal, unaccredited, voluntary, lifelong learning program participated in this study. The older adult students who participated had at least an elementary school education; some studied at the senior high school level, although not all attained a diploma and none had a college degree. Data were collected from an urban university in Taiwan. Before the study, the program was reviewed and approved by the university's institutional review board (ethics committee). Of the 208 respondents, 65 (31.2%) were traditional college students (aged between 18 and 22 years) and 143 (68.7%) were older adult students (aged from 55 to 72 years).

The answers to the questionnaire were evaluated to select prospective respondents as the second phase interviewees. The participants were chosen from a voluntary sample (see Table 1). This resulted in the selection of 5 traditional college students and 5 older adult students as the interviewees. Interviews took place from January 2017 to February 2017 and lasted for about 1 hour.

Table 1. Demographic data on the interviewees.

Interviewees, gender	Age (years)	Educational level
Traditional college student		
Male	20	Studying at university (sophomore, the second year of college)
Male	20	Studying at university (sophomore, the second year of college)
Male	19	Studying at university (freshman, the first year of college)
Female	21	Studying at university (junior, the third year of college)
Female	20	Studying at university (sophomore, the second year of college)
Older adult student		
Male	57	Senior high school
Male	62	Junior high school
Female	70	Junior high school
Female	59	Senior high school
Female	68	Elementary school

Instrument

The eHLS measures a student's ability to seek, find, understand, and evaluate health information from electronic sources and to apply this knowledge to address or solve a health problem [4,9,14]. A "gold standard" eHealth literacy instrument, which health researchers have recommended as an account of the social nature of eHealth, is being discussed by health professionals at the time of writing of this manuscript [23]. The 12-item eHLS is an instrument used to measure eHealth literacy among adult Taiwanese individuals. It was developed by Chiang et al [6],

who surveyed a representative sample of college students between March 2014 and May 2014, obtaining 455 valid responses. The reliability of the individual eHLS items ranged from .36 to .74. Standardized factor loading ranged from .60 to .86 (P<.001). Composite reliability ranged from .75 to .84, and the average variance extracted for each dimension ranged from .50 to .52. The indicators demonstrated a good fit for the measurement model. The scale includes functional (3 items, Cronbach alpha=.70), interactive (4 items, Cronbach alpha=.77), and critical (5 items, Cronbach alpha=.83) eHealth literacy dimensions. Its internal structure and external validity are considered acceptable. Respondents were asked to select the most accurate answer to describe their eHealth literacy level on a 5-point Likert scale, ranging from 1=strongly disagree to 5=strongly agree. Cronbach alpha of the overall scale was .84 (Multimedia Appendix 1).

One of the aims of this study was to try to capture the experience of the use of online health information for college students and older adult students. To do so, this study examined the participants' health literacy levels based on Nutbeam's [5] 3 levels of health literacy and definitions of eHealth literacy from Hsu et al (2011) [9] and Hsu et al (2014) [8]. Semistructured interviews were conducted to collect data. The interview outline included 4 items (see Table 2).

Data Analysis

Descriptive statistics and t tests were conducted to understand the effect of age on eHealth literacy. In addition, in conducting qualitative data analysis [24], the researcher conceptualized and developed a protocol to ensure open coding. Next, the analysis applied the concept of higher extraction level to "axial coding," which arranges concepts of similar content together into a class. The axial coding was then used, through classification, comparison, and induction, to analyze the subcategory and the main category together. To facilitate the classification and analysis of research data, the first column of Table 3 identified the participants, that is, A (traditional college student) and B (older university student). The second column identified the interview number. The third column identified an encoded serial number (eg, A-1-2).

Table 2. Interview guide items.

Focus point, open questions	The concept of health literacy
Access to online health information [7]	
What kind of online health information are you more interested in?	Basic ability to read health information (*functional literacy*)
What kind of source for online health information are you more interested in?	Knowledge of a supportive, interactive environment that provides health information (*interactive literacy*)
Literacy experience of online health information [3]	
How would you evaluate your internet health information reading and implementation experiences?	Basic ability to read and use health information (*functional* literacy and *interactive literacy*)
How do you assess the accuracy of health information? What is the assessment principle?	Both criticism of health information and its application to one's health *and* the response to that criticism, resulting in better control over living conditions (*critical literacy*)

Table 3. Coding and categorization examples.

Main category, subcategories, axial coding	Open coding
The experience of participants in obtaining online health information differed by group	
Access devices	
College students access most online health information from Web pages [A-2-64]	• We college students retrieved most online health information from Web pages. [A-5-31] • The most commonly used Web pages or websites are the occasional Yahoo News health section. [A-2-64]
Older adult students often obtained online health information through communication software [B-5-11]	• We have a Line group in the class; every day we share messages. [B-5-11] • Computers are used less now, as mobile phones are the most convenient. We often use communication software to share health information. [B-4-15]
Access types of online health information	
Traditional college students	• At University, my friends always pay special attention to their appearance, so I pay special attention to my looks. [A-1-28]
Older adult students	• Since retirement, I pay special attention to diet and health issues such as exercise and fitness It is good for health. [B-3-5]

The interview addressed the issue of participants' experiences in accessing online health information and focused on the concept of health literacy. In the process of data analysis, an academic peer was invited to use the code and test its relevance to meet the consistency and reliability requirements of qualitative research.

The purpose of this study was to understand the ways that subjects perceive their experiences. Therefore, the researcher played the role of "listener" in the interview process and did not make value judgments while conducting the interviews. Coding and categorization examples are shown in Table 3.

Results

Study 1—Quantitative Analysis of Electronic Health Literacy: Electronic Health Literacy Between Age Groups

There was a significant difference in the overall eHealth literacy scores between traditional college students and older adult students (t_{207}=2.98; P=.001), with the overall scores of traditional college students (mean 43.78) being higher than those of the older adults (mean 40.93). In the functional eHealth literacy dimension (t_{207}=12.17; P<.001), the traditional college students' scores (mean 11.43) were higher than those of the older adults (mean 8.08). However, there was no significant difference between interactive and critical eHealth literacy (see Table 4). In addition, in this study, we found a difference between age groups; functional eHealth literacy was higher than interactive and critical literacies for traditional college students (F=101.28; P<.001) and older adult students (F=373.24; P<.001).

Study 2—Qualitative Analysis of Electronic Health Literacy: Participants' Experiences and Performance of Reading Online Health Information

First Result: The Experience of Interviewees in Obtaining Online Health Information Differed Between Generations

The initial motivation for college students to access online health information was to meet the needs of beauty, weight loss, fitness, and so on:

> *Before in high school, I had a face full of acne and now have pockmarks. Now, at University, my friends always pay special attention to their appearance, so I pay special attention to my beauty.* [A-5-31]

> *Girls always want to have a good body shape...You may see someone post on the Internet about how they succeeded in losing weight. When we see a successful experience in which someone loses weight quickly, no matter whether it is true or not, we want to learn from them.* [A-3-28]

Table 4. Electronic Heath Literacy Scale means, SDs, and *t* tests.

Factor	Traditional college students (N=65), mean (SD)	Older adult students' (N=143), mean (SD)	The assumption of equal variances	*P* value	*t* test	*P* value
Functional electronic health literacy	11.43 (1.94)	8.08 (1.56)	3.706	.06	12.17 (206)	<.001
Interactive electronic health literacy	14.50 (2.67)	14.60 (2.52)	0.100	.75	−0.263 (206)	.79
Critical electronic health literacy	17.81 (3.74)	18.18 (3.28)	3.214	.08	−0.71 (206)	.48
Electronic health literacy	43.78 (6.68)	40.93 (5.10)	3.711	.06	2.98 (206)	<.001

Furthermore, the online health information retrieval methods of college students can be divided into *fixed* and *nonfixed* Web pages, which were used for 2 different purposes. Fixed habitual behavior of college students was to retrieve health information from static Web pages or magazines and either browse health knowledge or seek suitable skin care products for themselves, rather than solving practical health problems. On the other hand, they also gathered health information from nonfixed pipelines with the aim of solving health problems that were of immediate concern. Often, when respondents were aware of health problems, they would conduct online health information retrieval, looking up information about, for example, acne, weight loss, treatment of colds, gastrointestinal care, medical topics (such as new flu prevention), cancer diet, and so on. Data revealed that respondents searched using search engines and entered keywords but did not use search techniques involving, for example, Boolean logic. Respondents often looked only at the first search result. If there were many websites retrieved, there was a greater possibility that respondents would find a Web page with an appropriate answer or a Web page that was familiar to them. If there were links to other Web pages, they would click those links. However, if the health message of the interviewee was based on personal experience, there were limitations to his or her search for this kind of individual-oriented experience. Such internet health information was offered only as a reference, and it had little influence on actual implementation:

> *The most commonly used site is the occasional Yahoo news health section, on things like healthcare, healthy food, health exercises, and clicking on links to help with acupuncture points and the like, such as ah-shi acupuncture points for the eye.* [A-2-64]

> *I often use Yahoo or Google and get many results from forums...just by looking at a forum or seeing someone share an experience. Message-board posts*

> *can only be used as reference points that are not easy for individuals to adopt. This is because everyone has a different living environment and because the retrieved information cannot be used for all situations.* [A-4-73]

Respondents were most interested in reading online health information. When browsing health information, most of the information available online was not sufficiently clear or too specialized. Moreover, most respondents asserted that they understood the health information on the internet, but some proper nouns, foreign language terms, and various ways in which data were presented made it difficult for some respondents to understand and learn from adverse outcomes:

> *Most of the online health information is easy to understand and does not appear [for example] like a book that is written very professionally and difficult to read. Getting information from the Internet is very convenient, [because] you can quickly know the information and it will not be difficult [to understand].* [A-1-41]

Furthermore, the criteria for assessing the quality of online health information can be divided into 2 categories: subjective judgment and objective form. *Subjective judgment* is based on the cognitive judgment of the respondents; thus, they use their preknowledge to evaluate health information. Some would take the initiative by seeking others' advice to confirm the quality of online information, specifically by cross-validating through different Web pages or asking professionals (eg, doctors, pharmacists, medical friends, and relatives) directly to arrive at a more credible conclusion:

> *There is some very obviously illogical [information that] I will not go to, at least with the mind to question it. For example, burn soy sauce, or wipe pepper on your skin for weight loss, but you would think that this will only hurt your skin.* [A-5-35]

On the other hand, the *objective form* refers to the quality assessment of respondents from the external mode of online health information. In addition, some of the respondents would direct their attention toward the sources used by the specific internet source, particularly focusing on "update time," "certification mark," "whether it was an official website," "the number of visitors," "whether there are small ads," and so on:

> *Reading numbers, small ads...If the site appears with too many small ads, that is, with commercial activity, [with an intent] to sell some things, I will not accept it.* [A-3-98]

> *I will check information from an accredited institution...[If called to choose] [b]etween information available for 2005 and 2009, I would rather believe information from 2009, because it is relatively new. The content is really a relatively large problem; I will pay attention to the source...If it is just the Central Research Institute meeting with a doctor, I will ask that day to see that...that perspective?* [A-2-91]

Online health information used by older adult students was mostly concerned with "diet and nutrition," "health and wellness," and "exercise and fitness":

> *Before, when I was working, I was often busy when I ate, so I did not pay attention to nutrition, had no time to exercise, causing me to be physically ill often, and also susceptible to catching a cold, and now, after retirement, I pay special attention to diet and health issues such as exercise and fitness. I think exercise [is important] for your health.* [B-3-5]

Besides, older adult participants had access to a communications software group, enabling them to access online health information via group sharing:

> *We have a Line group in the class, [and] every day we share messages. Last month a student was sick and suddenly died. We are already into old age, so we all attach great importance to health problems, and if we do not maintain our health, we will soon meet God. So, we have a bunch of health messages every day to share with each other, every day mobile phone messages to forward.* [B-5-11]

> *I am now sixty years old, at the age when I begin to face what old, sick, dead, and living must go through, when you hear friends around who get sick and then survive, they become more self-alert. Computers are used less now, as mobile phones are most convenient, I now have time to draw...We often use communication software to share health information to friends and relatives, in the hope that people can keep healthy.* [B-4-15]

The above data indicated that the majority of older adult students had retired from work and had more time to pay attention to their health, focusing more on their "diet and nutrition," "health and wellness," and "exercise and fitness." Their primary source of information was mobile communication software through which they shared health messages.

Second Result

Most of the respondents believed that they had no problem with reading comprehension. However, most of the older adult students observed that they had lower critical ability and difficulty distinguishing correct information from incorrect information. Therefore, they believed that most health information was not very reliable as a reference:

> *There are a lot of opportunities to share health information, [and] reading comprehension is not a problem. I am also interested in the content, but some of the information overlaps, and I cannot evaluate the accuracy, so the reference value is reduced. Also, I have to follow the practice of health information. But the effect [once the information is applied] is not as favorable as the health information itself, so for these messages, after reading the reference, the effect is more difficult to control.* [B-3-45]

Older adult students primarily based their assessment of information accuracy on both subjective judgment (such as individual experience or previous knowledge) and objective

standards (such as sources, publication date, or professional authorship). These respondents used implicit perceptions of their own experiences and subjective judgments as well as explicit perceptions, such as information from experts and cross-validation of information from credible institutions:

> *Online health information needs to be cross-verified for correctness through different channels, such as seeking formal medical information medical information or asking a medical professional.* [B-1-71]

Discussion

Electronic Health Literacy Performances Vary Across Generations

In the overall assessment of eHealth literacy, traditional college students scored higher than older adult students. The results indicate that eHealth literacy varies across generations. eHealth literacy has multiple levels [6]: the first layer, *functionality*, is the basic reading ability; the second layer, *interaction*, is advanced knowledge; and the third layer, *criticality*, is more in-depth analytical ability to judge [25]. This may be explained by the concept of cognitive dimensions, as both functional and critical eHealth literacy involve higher levels of ability that require more extended periods of cognitive training to develop. In this study, young college students were more educated than older adult students. Therefore, it was likely that the level of educational attainment for these college students would be high, and, consequently, eHealth literacy levels would also be high [26]. Accordingly, college students had higher scores in eHealth than older adult students.

Individual Experience of Accessing Online Health Information

This study found that traditional college students and older adult students accessed online health information differently. Young adult college students accessed online information from websites and focused mostly on “beauty,” “weight loss,” and “fitness.” These topics typically concern physical appearance. Older adult students accessed online health information generally through communication software, enabling them to share more health information. The health information older adult students tended to access included information about diet and nutrition, health and wellness, and exercise and fitness.

Moreover, most of the study participants had expressed the belief that the majority of online health information could be read and understood. However, when evaluating the quality of information, they were generally doubtful. Participants pointed out that the quality of online health information is divided into subjective judgment and standard objective information. However, some participants in each group suggested that, in the era of information explosion, it is not easy to choose and determine the accuracy of the information. Thus, the participants tended to take a skeptical attitude if they were personally involved and would then find additional resources to verify through multiparty comparison to enrich their knowledge and understanding of the topic. This finding is similar to the findings in the study by Hsu et al [8]. This may be explained by the Uncertainty Management Theory (UMT) [27]. A prominent communication uncertainty framework has been applied to appraise the associations between online health information seeking and uncertainty management [28,29]. A central tenet of UMT proposes that uncertainty is not necessarily a negative or positive experience; instead, an individual will appraise the meaning of uncertainty, and the resulting emotional response will determine whether the uncertainty is evaluated as negative, positive, or neutral. The uncertainty evaluation will influence an individual’s behaviors in managing his or her uncertainty. For example, individuals for whom uncertainty is an undesirable or negative state may seek health information to augment their knowledge and thereby lessen their state of uncertainty [30].

Furthermore, in study 1, functional eHealth literacy scores were different for traditional college students and older university students, but interactive and critical eHealth literacy scores did not differ. Empirical evidence has shown that an inverse correlation exists between age and eHealth literacy [17,31,32]. In study 2, the experience of how interviewees obtained online health information differed according to generation. Most interviewees could read online health information, but older adult students had less critical ability to do so. Paige et al [14] indicated that, compared with younger adults, older adults had less ability, possibly owing to cognitive degradation, to evaluate and act upon online health information. Bodie and Dutta [2] pointed out that antecedents such as personality and educational background are the factors that influence the individual’s eHealth literacy. The interviewees in this study received different modes of cognitive training, which might be responsible for the difference in the results and also highlights the need for eHealth literacy to be introduced into education.

Limitations

This study has several limitations. First, the study was performed under trial conditions in which it was expected that the participants would answer hypothetical questions that were not clear and did not directly address the participants’ actual physical health concerns. Second, this study found that both younger college students and older adult students were capable of reading and understanding online health information to some extent. The participants did have basic eHealth literacy. However, both groups experienced difficulties in recognizing correct health information online. Given that we did not observe the highest level of critical eHealth literacy, the data collected through our investigation may not have been specific enough. Furthermore, the interviewees constituted a voluntary sample of 5 participants, which likely affected the generalizability of the results.

This study found that functional eHealth knowledge varies across age groups, unlike interactive and critical scores. Higher levels of eHealth knowledge are more difficult to cultivate and evaluate [5,6,8]. Given that most people do not have the ability to attain higher-level health literacy, there were no differences between groups. However, some studies have pointed out that older adults had less confidence in eHealth resource awareness, information-seeking skills, and the ability to evaluate and act upon online health information [14]. This study used interviews to understand the participants’ capacity to read, understand, and

critically evaluate health information, along with other relevant experience. It collected and analyzed the participants' self-reported data. However, there is a difference between the online performance of the participants and their real offline experience of eHealth literacies. Future studies could record the actual online behavior of the participants, utilizing a specific observable behavior analysis, and investigate how the subjects demonstrate the 3 levels of eHealth literacy. This would give a better estimation of their actual abilities [26].

Conclusions

This study explored the effect of age on eHealth literacy in the hope that the findings will stimulate further debate about how a health education framework can be translated into practical approaches and contribute to the further refinement of the eHealth literacy concept. In the eHealth literature, this is the first study to explore whether group differences exist.

The findings showed that there are gaps in eHealth literacy between traditional college students and older adult students. Therefore, we recommend that the education system strengthens the functional and critical eHealth literacy of both groups. These age disparities have been attributed to older adults' unique health needs (eg, "diet and nutrition," "health and wellness," and "exercise and fitness") as compared with younger adults [33] and to more specialized health concerns (eg, beauty, weight loss, and fitness) of college students.

Due to different personal eHealth literacy experiences, to develop specific strategies for promoting individualized health literacy, it is necessary to know about the current health concerns of different individuals not only to design eHealth learning programs but also to empower individuals in planning them. The study observed that participants, when faced with challenging and uncertain health situations, employed various strategies to reduce ambiguity about a health-related condition. Furthermore, based on Nutbeam's original conceptualization, Paige et al [34] proposed the Transactional Model of eHealth Literacy, which has theoretical underpinnings in transactional communication literature, and adds a fourth level of "translational eHealth literacy." This is the highest cognitive level of eHealth literacy, and it is informed and builds upon lower-level eHealth literacy dimensions (ie, critical, communicative, and functional). Future studies could investigate how the internet may provide a useful and valuable channel for health information to consumers who wish to utilize information strategies for managing health-related uncertainty.

Moreover, literacy, numeracy, decision-making, and reasoning skills may be needed for the critical evaluation of the retrieved information by each group of students [26]. These skills reflect more critical and translational elements of online information processing, inviting future study. The study summarizes online health information experiences and performances of traditional college students and older adult students. As data collection may involve participants' subjective perception of their abilities, it is not easy to determine skill-based literacies. Future research is needed to explore this issue.

Acknowledgments

This research is supported by the Ministry of Science and Technology of Taiwan, project number MOST 106-2511-S-151-003.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The eHealth Literacy Scale.

[PDF File (Adobe PDF File), 15KB - humanfactors_v6i2e11480_app1.pdf]

References

1. Taiwan Network Information Center (2017). 2017 Taiwan Broadband Internet Usage Survey Report URL: https://www.twnic.net.tw/download/200307/20170721e.pdf [accessed 2019-04-10] [WebCite Cache ID 77Ww6eTbF]
2. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jul;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
3. Eng T. The e-Health Landscape: A Terrain Map of Emerging Information and Communication Technologies in Health and Health Care. New Jersey: The Robert Wood Johnson Foundation; 2001.
4. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
5. Nutbeam D. Literacies across the lifespan: health literacy. Lit Numer Stud 1999;9(2):55 [FREE Full text]
6. Chiang C, Yang SC, Hsu WC. Development and validation of the e-health literacy scale and investigation of the relationships between E-health literacy and healthy behavior among undergraduate students in Taiwan. Formosa J Ment Health 2015;28(3):420. [doi: 10.30074/CJMH]
7. Werts N, Hutton-Rogers L. Barriers to achieving e-health literacy. Am J Health Sci 2013 Aug 14;4(3):115. [doi: 10.19030/ajhs.v4i3.8007]

8. Hsu WC, Chen SF, Ho CJ. Experience of using web health information among college students: an analysis from the health literacy perspective. J Health Promot Health Educ 2011;35:22. [doi: 10.7022/JHPHE.201106.0001]
9. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
10. Han J. China Credit Information Service. 2009. Internet penetration rate among countries URL: http://www.credit.com.tw/newweb/market/weekly/index.cfm?sn=46 [accessed 2019-04-11] [WebCite Cache ID 77ZGSMjRv]
11. Yeh CT. The managers' experiential learning of program planning in active ageing learning centers. J Educ Pract Res 2016;29(2):105-136.
12. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
13. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
14. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: measurement invariance study. J Med Internet Res 2018 Jul 09;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
15. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
16. van Deursen AJ, van Dijk JA. Using the internet: skill related problems in users' online behavior. Interact Comput 2009 Dec;21(5-6):393-402. [doi: 10.1016/j.intcom.2009.06.005]
17. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
18. Liou MY, Lu CM. Related factors of health behavior in university students. Chinese J School Health 2006;48:19-37. [doi: 10.30026/CJSH.200606.0002]
19. National Development Council. 2018. Aging time in Taiwan URL: https://www.ndc.gov.tw/Content_List.aspx?n=695E69E28C6AC7F3 [accessed 2019-04-11] [WebCite Cache ID 77ZGydCRN]
20. Liu C, Liao L, Shih SF, Chang T, Chi H, Osborne R. Development and implementation of Taiwan's child Health Literacy Test. Taiwan Journal of Public Health 2014;33(3):251-270. [doi: 10.6288/TJPH201433102105]
21. Su CL, Chang SF, Chen RC, Pan FC, Chen CH, Liu WW. A preliminary study of Taiwan Health Literacy Scale (THLS). J Formos Med Assoc 2008;12(5):525-536.
22. Venkatesh V, Brown SA, Bala H. Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. MIS Q 2013;37(1):54 [FREE Full text]
23. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
24. Chen SM. Qualitative research in social science. Taipei: Wu-Nan Book Inc; 2002.
25. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
26. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Human Behav 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
27. Brashers DE, Neidig JL, Haas SM, Dobbs LK, Cardillo LW, Russell JA. Communication in the management of uncertainty: the case of persons living with HIV or AIDS. Commun Monogr 2000 Mar;67(1):63-84. [doi: 10.1080/03637750009376495]
28. Rains SA, Tukachinsky R. An examination of the relationships among uncertainty, appraisal, and information-seeking behavior proposed in uncertainty management theory. Health Commun 2015 Jun;30(4):339-349. [doi: 10.1080/10410236.2013.858285] [Medline: 24905910]
29. Rains SA, Tukachinsky R. Information seeking in uncertainty management theory: exposure to information about medical uncertainty and information-processing orientation as predictors of uncertainty management success. J Health Commun 2015 Jul;20(11):1275-1286. [doi: 10.1080/10810730.2015.1018641] [Medline: 26132807]
30. Stone AM, Scott AM, Martin SC, Brashers DE. Using information to manage uncertainty during organ transplantation. Qual Commun Res 2013 May;2(1):42-60. [doi: 10.1525/qcr.2013.2.1.42]
31. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
32. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
33. Bennett JA, Flaherty-Robb M. Issues affecting the health of older citizens: meeting the challenge. Online J Issues Nurs 2003;8(2):2. [Medline: 12795628]

34. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]

Abbreviations

eHealth: electronic health
eHLS: electronic Health Literacy Scale
UMT: Uncertainty Management Theory

Edited by G Eysenbach; submitted 03.07.18; peer-reviewed by S Paige, J Miller, B Greer; comments to author 08.10.18; revised version received 02.12.18; accepted 29.03.19; published 21.04.19.

Please cite as:
Hsu WC
The Effect of Age on Electronic Health Literacy: Mixed-Method Study
JMIR Hum Factors 2019;6(2):e11480
URL: http://humanfactors.jmir.org/2019/2/e11480/
doi:10.2196/11480
PMID:31066696

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org