JMIR Human Factors

Impact Factor (2022): 2.7
Volume 3 (2016), Issue 1 ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

Contents

Viewpoints

Editorial

Original Paper

Using the NIATx Model to Implement User-Centered Design of Technology for Older Adults

David H Gustafson Jr[1], BA, MS; Adam Maus[1], BS, MS; Julianne Judkins[1], BS; Susan Dinauer[1], BA; Andrew Isham[1], BA, MS; Roberta Johnson[1], BA, MA, MEd; Gina Landucci[1], BS; Amy K Atwood[1], BA, MS, PhD

Center for Health Enhancement Systems Studies, Department of Industrial and Systems Engineering, University of Wisconsin - Madison, Madison, WI, United States

Corresponding Author:
David H Gustafson Jr, BA, MS
Center for Health Enhancement Systems Studies
Department of Industrial and Systems Engineering
University of Wisconsin - Madison
4145 Mechanical Engineering Building
1513 University Avenue
Madison, WI, 53706
United States
Phone: 1 608 262 1746
Fax: 1 608 890 1438
Email: dgustafson@wisc.edu

Abstract

What models can effectively guide the creation of eHealth and mHealth technologies? This paper describes the use of the NIATx model as a framework for the user-centered design of a new technology for older adults. The NIATx model is a simple framework of process improvement based on the following principles derived from an analysis of decades of research from various industries about why some projects fail and others succeed: (1) Understand and involve the customer; (2) fix key problems; (3) pick an influential change leader; (4) get ideas from outside the field; (5) use rapid-cycle testing. This paper describes the use of these principles in technology development, the strengths and challenges of using this approach in this context, and lessons learned from the process. Overall, the NIATx model enabled us to produce a user-focused technology that the anecdotal evidence available so far suggests is engaging and useful to older adults. The first and fourth principles were especially important in developing the technology; the fourth proved the most challenging to use.

(JMIR Human Factors 2016;3(1):e2) doi:10.2196/humanfactors.4853

KEYWORDS

eHealth; user-centered design; technology; aging in place; independent living; consumer participation; accessibility

Introduction

In 2010, the Federal Agency for Healthcare Research and Quality started funding the Active Aging Research Center (AARC) to develop technology to help older adults live longer independently [1]. The AARC is housed at the Center for Health Enhancement Systems Studies (CHESS) at the University of Wisconsin-Madison. CHESS has been building and testing information and communication technologies (ICTs) for patients and their families since the 1970s. Previous CHESS ICTs have been proven effective in numerous randomized trials for a variety of conditions, including alcohol use disorders [2], lung cancer [3,4], pediatric asthma [5], breast cancer [6], and HIV [7].

In the original AARC grant application, we defined what the technology for older adults would accomplish, building on previous CHESS systems, but did not identify a specific technological solution. Instead, we planned to develop a technology for adults aged 65 and over by working closely with older adults themselves as well as with informal caregivers, health care professionals, community members, and others; test the technology in a randomized controlled trial [8]; and, if the technology proved to be effective, disseminate it.

One assumption we had when we started this work was that older adults are rarely the target of technology development. Although guidelines exist for designing technology for older adults [9] and the literature reports some efforts to develop technology with and for older adults [10,11], we found few

easy-to-use websites and interfaces designed specifically for older adults. Yet recent research shows a significant increase in technology use among older adults. Between 2008 and 2012, adults aged over 65 had an increase of 39% in Internet use, the largest among all age groups, and now 50% of all older adults are online [12]. Older adults are willing to use technology if they think it adds value and convenience to their lives and supports their activities [13]. While older adults have physical limitations that can make using technology challenging, such as low vision, dexterity problems, and cognitive issues [14], we believed that designing a technology with and for older adults would help overcome barriers to use [15].

This paper reports on using a customer-focused process improvement model (Network for the Improvement of Addiction Treatment [NIATx]) as the user-centered design (UCD) approach to developing technology for older adults.

UCD at CHESS

Central to UCD is the principle that having a thorough understanding of the end user's needs and capabilities is essential to creating the most effective system or product [16]. Since the late 1980s, when early publications [17] sparked an interest in applying UCD to technology development, various methods and models of UCD have been extensively researched [18]. Despite the work done on UCD research and application, UCD remains loosely and variously defined [18].

At CHESS, the tech team used UCD methods (usability testing, card sorting, paper prototyping, focus groups, surveys, etc [19]) without having a structure for using these methods at different stages of development. In addition, we often ran out of time or lacked the resources to implement UCD methods throughout a project. As UCD expert Jakob Nielsen [20] pointed out, many developers abandon UCD methods because of cost, time, and complexity, and this was the case at CHESS. Although we agree with Karat [21] that UCD does not need to be a rigid set of practices, we sought practical key principles that would provide a structure for the application of UCD methods throughout the life cycle of product development. Gulliksen et al [22] defined 12 key principles of UCD based on standards and experience in using various models in a variety of projects. Their work also includes lists of activities that relate to each principle. Even with these well thought out and researched principles, we became overwhelmed with the options and activities that could be used and lacked the time and resources to research alternative models of UCD.

The tech team at CHESS is relatively small, with 2 software developers; 1 user-interface designer; 1 Web master; 2 information technology professionals supporting hardware, infrastructure, and the helpline; and the tech director. Each tech project usually involves a manager, a software developer, and a user-interface designer. In addition, tech team members are a shared resource at CHESS, meaning that individuals usually work on multiple projects at one time. Because of the size of the AARC grant and the number of different goals, academic departments, and principal investigators (PIs) involved, we felt we needed clear guidelines on how to apply UCD within the development process, so that we could incorporate user feedback in design decisions, promote speed, and keep team members informed about progress. Table 1 shows the organization of the AARC project. The technology developed during the project ultimately became a website called "Elder Tree."

Table 1. Organization of the Active Aging Research Center project.

Center individual or sub-group	Main functions
Lead principal investigator (PI)	Generates ideas, overall management and priorities, final decision making. The lead PI is the Director of Center for Health Enhancement Systems Studies, where the Active Aging Research Center is housed.
Project director	Day-to-day management of overall project
Research teams	Each team works on 1 of the following 5 challenges for older adults: isolation and loneliness, driving and transportation, caregiving, medication management, and falls prevention. Each team has a PI, change leader, and team members.
Community partners	Identify the needs and assets of older adults; provide feedback on evolving iterations of the Elder Tree technology. Community partners are older adults, the Wisconsin Institute for Health Aging, and local Aging and Disability Resource Centers.
National Advisory Committee	Review of plans and progress, advice on Elder Tree technology and research. The committee consists of 17 nationally recognized advisers in gerontology, technology, public policy, medicine, communications, driving and highway safety, organizational change, and other areas.
Tech team	Design and development of information and communication technologies for patients and family members, including Elder Tree.
County coordinators	Local management of the randomized trial (recruitment, training, etc). County coordinators are grant-funded employees, 1 in each of the 3 regions where Elder Tree is being tested.
Strategy teams	Through interviews, identify needs and assets of older adults in each community. Strategy teams consisted of citizens from each of the 3 regions where Elder Tree is being tested.

The NIATx Model

The NIATx model, developed at CHESS, is a simple framework for process improvement. NIATx originally stood for "Network for the Improvement of Addiction Treatment"; the NIATx model was initially used to improve retention and access to care in behavioral health agencies [23]. Now NIATx functions as a word referring to a division of CHESS that teaches and conducts

process improvement in a range of health care and other organizations.

The model, which has only 5 principles, was intended to be easy to learn and implement so that individuals with little or no knowledge of process improvement can quickly test changes to improve services and outcomes. The model has no levels of training or complicated data elements to collect. Although not specific to technology development, the model is a user-centered approach intended to be applied flexibly [23,24]. The NIATx model is evidence based, and many of us on the tech team were familiar with it from developing tools to support NIATx research projects conducted at CHESS. Unlike other models of UCD, the NIATx model includes a method for developing innovative solutions. For these reasons—ease of use, flexibility, familiarity with the model, the model's evidence base, and its approach for developing innovative solutions—we decided to apply the NIATx model as a UCD framework that would give structure to the development process.

The Five Principles of the NIATx Model

The NIATx model rests on 5 principles that have been shown to be the essential elements of successful change projects [25]. These principles were developed from analyzing decades of research from 13 different industries related to why some projects fail and others succeed [25].

As we considered using the NIATx model as a framework for UCD, 3 of the principles—rapid-cycle testing, understanding and involving the customer (or end user), and getting ideas from outside the field—seemed especially useful because they would enable us to test innovative ideas quickly, assess their effectiveness, incorporate user feedback, and make additional changes rapidly.

The first principle, *understand and involve the customer,* is the most important principle of the NIATx model. In fact, this principle has more impact than all the other principles combined [25]. "Customer" refers to the end user, who may or may not pay for the product or service being designed or implemented. Allocating the time and resources to deeply understand the needs and assets of end users and getting regular ongoing end user feedback increase the likelihood that a product will succeed. In this paper, we refer to the customer as the "end user" or "older adult."

The second principle, *fix key problems*, arises from the understanding that a project is more likely to succeed if top management is involved and committed to the project, in part because this makes it more likely that the support and resources needed to succeed will be available. One way to ensure the commitment of top management is to address the key issues or problems that top managers face. One strategy is to ask, "What keeps the CEO awake at night?" The answer identifies the problem(s) to start with.

The third principle is to *pick an influential change leader*. The role of the change leader is to move a project forward by identifying and removing barriers to progress. An influential change leader is a staff member who has respect from management and staff, is a good leader, and has direct access to the CEO and other critical stakeholders. The change leader should have the authority to do whatever it takes to keep a project moving forward.

The fourth principle, *get ideas from outside the field,* is the second most important NIATx principle [25]. It can be broken down into 3 phases. First, identify a field or fields that face problems similar to the problems your organization faces. Second, find the organization in that field that is best at dealing with that problem. Third, identify what makes that organization so much better than others at addressing that problem. This process forces you to identify the core problem you are facing and can lead to innovative solutions. Atul Gawande [26] used this principle when he described the possible application of coaching, as done in sports and music, to the work of surgeons and the application of cost and quality control, as done in restaurant chains, to health care [27]. Looking outside the field of UCD for an approach to technological development, as we did in this case, can be considered another example of this principle.

As an example of using this principle, staff members who wanted to improve teamwork in their organization would begin by asking, "What industry requires good teamwork?" One answer might be National Association for Stock Car Auto Racing (NASCAR); the pit crew of a NASCAR team demonstrates exceptionally good teamwork. One particularly good pit crew works for NASCAR driver Denny Hamlin [28]. Staff members then identify the key characteristics of Hamlin's crew that make the crew successful. For instance, crew members work together seamlessly under very stressful conditions. Each member has a clearly defined job, and each understands every other member's job. Pit crews constantly practice to stay sharp for the race, and their performance is constantly measured. With these characteristics identified, staff members can now apply the ideas to their own work environment. Looking outside the field can produce solutions not previously considered.

The fifth principle, *use rapid-cycle testing*, encourages developers to develop small improvements and test them with end users and stakeholders to see how they work. After each test, the improvement (or in our case, tech) team makes changes and then tests again. Several cycles of rapid-cycle testing help create a high-quality product on release or an improvement in a process that actually works. Rapid-cycle testing was first described by Shewhart [29] and revised and popularized by Deming [30]. For example, if stakeholders identified reducing the home page bounce rate (the rate at which users abandon a website after landing on the first page) as a key problem to solve, usability tests would be used early in development to determine what seems to cause users to leave the home page. The answer(s) would determine changes to make in the home page, and the success of the changes would be determined by another wave of usability testing. The process would continue until stakeholders were satisfied that the bounce rate had been sufficiently reduced.

Applying the NIATx Model

Understand and Involve the Customer

Before starting to develop a technology to help older adults continue to live independently, we began the process of understanding our end user. What we learned drove development, including what services to provide in the technology, what content to include, and the overall design of the system [31]. The Elder Tree website in use in the randomized trial as of this writing is available in an archived version [32].

All staff members working on the AARC project (PIs, researchers, tech team members, administrative staff—see Table 1) were asked to take part in focus groups or one-on-one interviews with older adults and their caregivers. During the course of the project, hundreds of interviews and more than 20 focus groups were conducted. All staff were asked to write, for each individual they talked to, a story that summarized the older adult's experiences and current situation. Creating stories from these interviews brought the experiences of older adults to life, giving us a better understanding of and greater empathy with their day-to-day challenges and joys. The AARC staff members met regularly to share and discuss the stories. From defining high-level goals to conceptualizing and developing solutions, these stories were foundational to our development work.

During this period when all project staff members were getting to know firsthand the assets and needs of older adults, tech team members volunteered at a local senior center to better understand how older adults use and learn to use new technology. Members of the tech team designed and taught a series of classes on using computers and the Internet. Topics included Internet basics, Internet safety, Facebook, Skype, and downloading and managing digital images. While a tech team member led a class, other team members circulated among the students to offer one-on-one support. Seeing and experiencing, up close, how older adults interact with technology had a profound effect on our work. For example, we observed that arthritic hands had trouble using a mouse and that Web pages with many sections and subsections were hard for some older adults to understand and hard for others to see. These and many similar observations directly influenced the design of the Elder Tree website.

From all of these interactions with older adults, a few issues surfaced repeatedly. Older adults frequently expressed a concern about their safety on the Internet. They did not want to get scammed, lose money, or be asked to give private information. It also became clear we would need to address older adults' decreasing motor dexterity and problems with vision and hearing both in the computer we selected for older adults to use and the interface design. During the development of the technology, we often found ourselves rejecting accepted Internet conventions for the sake of accessibility. For example, we decided not to use rollover effects to display additional information. Users who struggle with a mouse can be distracted and disoriented by the rollover effects as they navigate a page. We also rejected a dashboard-type home page that would give a dense display of information. Instead, we embraced a simplified design with the goal of having a single task per page.

The process of understanding the end user produced another effect: It forged personal connections between tech team members and older adults and turned all of us tech team members into advocates for our end users.

Fix Key Problems

One way to identify key problems is to ask, "What keeps the CEO awake at night?" These problems are good ones to start with because top management will more likely be engaged in a project that addresses these problems and give the project the attention and resources it needs to be successful. Applying this principle to developing technology in our grant-funded project required some modification. The overall goal of the grant was to help older adults live longer independently. The PIs who applied for the grant took this as their mission; we regarded it as the answer to what was keeping the CEO up at night. However, this statement of the problem was too broad to suggest specific development steps, so we asked older adults themselves to help identify key problems more specifically.

In addition to the work done at the beginning of the project described earlier (conducting focus groups and interviews with older adults and summarizing the results in stories, and volunteering to teach older adults about the use of technology at a senior center), early work on the AARC project included a process called Asset-Based Community Development (ABCD) to learn about older adults in their communities and lay the foundation for dissemination [33]. Because our grant-funded project would culminate in testing whatever technology we developed in a randomized trial in 3 regions of Wisconsin (urban, suburban, and rural), we implemented ABCD in 1 community in each of the 3 regions. A staff member from CHESS with extensive experience in using ABCD worked with the county coordinator—someone who lives in the target area and was hired with grant funds to manage the project locally—to implement ABCD. The CHESS researcher and county coordinators led the creation of strategy teams made up of citizens from each community. These citizens came from local agencies, businesses, institutions, and organizations. Strategy team members and other volunteers interviewed friends and neighbors to create an inventory of the assets, challenges, and aspirations of older adults and their caregivers in the community. In all 3 regions, 80 home visits were conducted; many more focus groups and interviews took place in other settings. Remarkably, the top 3 problems for older adults were the same in each of the 3 communities: isolation and loneliness, not knowing about community resources and events, and transportation to and from resources and events.

Throughout the development work, we also relied on the expertise of researchers and community partners to inform our work (Table 1). Researchers came from the fields of falls prevention, geriatrics, driving, transportation, and innovation. A National Advisory Committee consisted of leading thinkers from these and other fields. This committee met annually to advise us on the progress and direction of the project. Community partners, researchers, and National Advisory Committee members identified other key problems affecting the ability of older adults to live independently, especially medication adherence, dementia, and depression.

Given all the problems identified by various stakeholders involved in the project, whose view of the key problems should take precedence? Were the various PIs, collectively, the CEO? Or were the community partners? What about potential payers for the final product of the study? If the technology does not address the problems of payers, they will not be willing to fund implementing the technology that results from this research. And what about older adults themselves? If we did not address their key problems, they would not use the system. All these groups had important ideas and insights that we needed to consider.

After looking at all this information, our lead PI decided that we should first develop and test something to address isolation and loneliness. This would allow us to address the top issue identified by older adults themselves; we tech team members also thought we could develop something quickly and rapidly test and retest it. AARC researchers would continue to conceptualize solutions for the other issues, such as falls prevention and medication management, but initially the tech team would focus on this particular issue.

Get Ideas From Outside the Field

Early in our development work, we focused on reducing the number of older adults being affected by scams because older adults so often raised this issue. We looked into examples of excellence in sales, lie detection, psychic cold reading, and cognitive interviewing, collecting ideas that might be repurposed. Cold reading, which is how a psychic creates the illusion of knowing someone he or she does not know, seemed to have the greatest potential to reduce scams. Cold reading uses several techniques (eg, fishing, vagueness, push statements, switches); for each, a block can be used. We began to discuss with older adults their use of the equivalent of blocks when they used technology, but abandoned this because older adults were so pervasively concerned about scams that we instead created a closed system that could be used only by vetted participants.

We also looked for ideas outside the field when we brainstormed topics to include in the discussion group. We asked ourselves to think of other ways older adults obtained information. One of our team members suggested newspapers. We decided to base our initial topics in the discussion group on the different sections of a newspaper (eg, sports, local news, arts, entertainment). Our premise was that this organizational system would feel familiar to most older adults.

Pick an Influential Change Leader

At CHESS, the change leader for a research project is responsible for driving a project forward and removing barriers to progress. Chosen by the CHESS director and co-director, the change leader understands and represents the needs of the end user, is committed to seeing the project succeed, has a strong belief in the value of the project, can defend his or her position articulately, and is someone other people can easily follow or find charismatic; the change leader has direct access to the PI of the study. A change leader is responsible for leading and motivating the team, organizing meetings, removing roadblocks, and delegating. Because designing and developing a new technology for older adults was a large effort involving research teams (1 team each for isolation and loneliness, driving and transportation, caregiving, and so on), a project director led the work of the research center overall, and a change leader emerged from the tech team to lead development work on each research team. Change leaders, working under the guidance of the project director, understood and represented the needs of the end user; acted as a liaison between the tech team and older adults, researchers, and other stakeholders; and removed barriers in the tech team's way.

Use Rapid-Cycle Testing

Quickness is the essence of rapid-cycle testing. Each test should take from a few hours up to a few weeks, depending on what is being tested [23]. Being clear about 2 things speeds testing: What are you trying to accomplish? How will you know whether the change is an improvement? Instead of taking months to design an entire system, a development team can create a piece of the system, quickly test it with users to get feedback, make changes, and retest. We used 2 methods to conduct rapid-cycle tests: (1) one-on-one usability testing and (2) pilot tests with older adults in the field. Usability testing helped us understand whether older adults could navigate the interface and allowed us to see what they did as they used it. We usually tested usability in 1 sitting with an older adult or adults, using paper prototypes or early builds of the system. Pilot tests took 2 weeks or more and involved older adults using various iterations of the system in their homes. Pilot tests allowed us to see whether users thought the system was helping them and were likely to keep using it.

Our initial development work focused on creating a simple online discussion group. We wanted to test whether older adults with limited or no experience with technology would be able to use the system. Once we had a working website, we conducted a single session of one-on-one usability tests at a local senior center. This session took just a few hours, with 3 participants taking part. From these tests we were able to identify glaring usability issues, such as how we were indicating clickable buttons.

After making changes, 10 older adult volunteers were recruited to take part in a 2-week pilot test. Participants were given a computer and access to the Internet through a mobile hotspot if they did not have their own Internet access. Participants were able to select from a number of devices, including 7-in. Android tablets, iPads, laptops, and 23-in. touchscreen all-in-one desktops. The AARC staff visited each participant at the participant's home; set up the computer and Internet connection, if necessary; and trained each participant on how to use Elder Tree. In all, we provided a computer and Internet connection to 9 of the 10 participants. During the 2-week pilot, AARC staff acted as discussion group "seeds," actively engaging with participants online to make sure Elder Tree had new messages and comments every day. After the 2-week pilot, each older adult was interviewed about his or her experience. This information, along with use and observational data from the pilot, was analyzed to make decisions about changes to make in the system. Work began on making those changes to run a second pilot test.

In total, 5 pilot tests were conducted involving more than 100 older adults. Eventually, we had enough older adults using the system that we were able to turn our discussion group into a beta version of the final website. Instead of running a pilot, getting feedback, making changes, and running another pilot test, we kept the site active for participants to use, solicited feedback from users in the discussion group, and rolled out improvements as they were completed. Using the discussion group for feedback proved to be an effective method of gauging the perceived usefulness and appeal of the site. The pilot tests and the beta site also supplied older adult volunteers who agreed to continue using the site during the randomized trial to act as "seeds" and peer mentors to study participants.

As we improved the site and developed new features, we continued to run ad hoc usability tests at senior centers and get ideas and feedback from researchers, experts on the National Advisory Committee, and our community partners.

Discussion

General Findings and User Views

Using the NIATx model has allowed us to develop a technology based on the needs and capabilities of end users. Anecdotal evidence suggests that using the technology is reducing isolation and loneliness among older adult users, which was the most important challenge older adults identified in the interviews, focus groups, and surveys we conducted. The views of some Elder Tree users are presented in Textbox 1.

Textbox 1. Views of some Elder Tree users.

- Since my husband died, I rarely get out of my house and Elder Tree has saved me.
- Elder Tree has given me back a sense of belonging.
- Elder Tree has connected me with people who are going through the same challenges as me. We are there for each other...I like that.

In addition, although we are still conducting the randomized control trial, early use data indicate that older adults facing barriers to technology adoption are using the website. Technology adoption generally is lower among those with less income, those with less experience using technology, and those with the least education [34]. Looking at study participants who created content on Elder Tree, meaning those who wrote or commented on discussion group messages, we found that 20.1% (27/135) created content 5 or more times a month (Table 2). Of this group, which we call super posters, 74% (20/27) did not have a computer or Internet access before the study, 19% (5/27) had a 4-year degree or above, and 89% (24/27) found dealing with finances challenging or difficult (Table 3).

Table 2. Categories of content creators (N=135).

Category	n (%)
Super posters (wrote >5 messages/month after training)	27 (20.0)
Medium posters (wrote ≥1 but <5 messages/month after training)	33 (24.4)
Low posters (wrote ≤1 message/month after training)	39 (28.8)
Did not post (Never wrote a message after training)	36 (26.6)

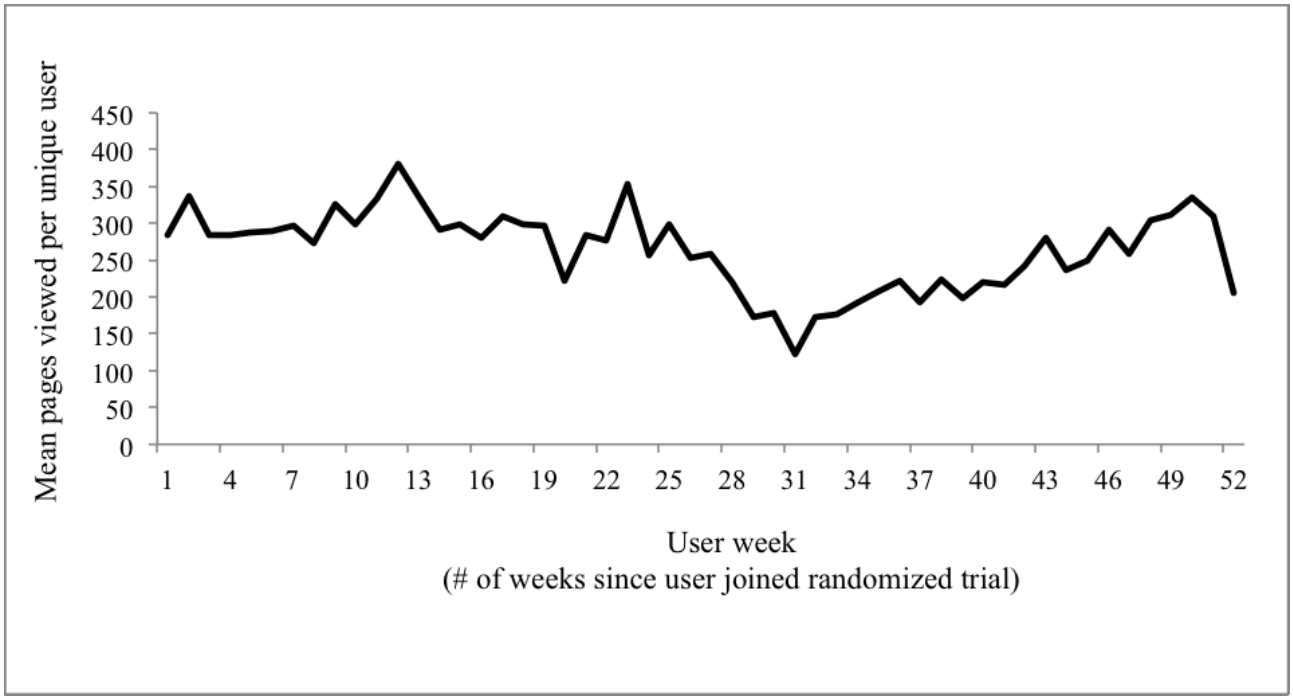
Table 3. Demographic characteristics of Elder Tree users.

Characteristic	Super posters (N=27) n (%)	Medium posters (N=33) n (%)	Low posters (N=39) n (%)	Did not post (N=36) n (%)
Did not have computer with Internet connection before the study.	20 (74)	15 (46)	22 (56)	16 (44)
Education (4-year degree or above)	5 (19)	8 (24)	8 (21)	11 (31)
Find dealing with finances challenging or difficult	24 (89)	22 (67)	19 (49)	17 (47)

We also looked at the overall use of the website by all participants, not just those creating content. When looking at mean pages viewed per user (Figure 1), we see a decrease in use after 6 months on study. However, near Month 8, we see a gradual increase to levels near the start of the study.

Although the qualitative and quantitative use data suggest that our development approach has led to a technology that older adults use, we also encountered problems and learned some lessons, which are described in the following sections.

Figure 1. Mean Elder Tree pages viewed per user.



Balance Efforts to Understand the End User Against Resources

Turning face-to-face meetings with end users into stories is a powerful exercise in understanding their needs and assets, but having everyone in a complex project (top management, tech team members, and research and administrative staff) conducting face-to-face interviews is very time consuming. We believe that the more people who see and hear from end users firsthand, the better, but we recognize that this approach might not be feasible in every organization. At the very least, 1 member of the development team, probably the change leader, needs to take the role of user advocate. This person should spend time in the field interviewing end users and conducting usability tests and be the voice of the user when conceptualizing new features and designs, therefore helping keep the project team focused on features that have the greatest efficacy.

Using a community-based process such as ABCD consumes considerable time and resources. Because we included this process in our grant application, we had the required resources for it. Although the process produced insights that helped define key problems, most organizations would not have the resources to use the process. We would rely in the future on focus groups; individual interviews with end users; creating personas to represent the different users of the technology, each with its own needs and assets; and the expertise of researchers and community partners to determine the problems faced by end users.

Beware of the Power of An Individual's Story

Storytelling is a powerful tool and brings the challenges of end users to life. However, individual stories can be almost too powerful. In our project, the moving story of 1 adult occasionally shut down what might otherwise have been a productive discussion of an improvement or new feature. Looking for common themes in multiple stories helps prevent a single story, or several, from having too much weight in development.

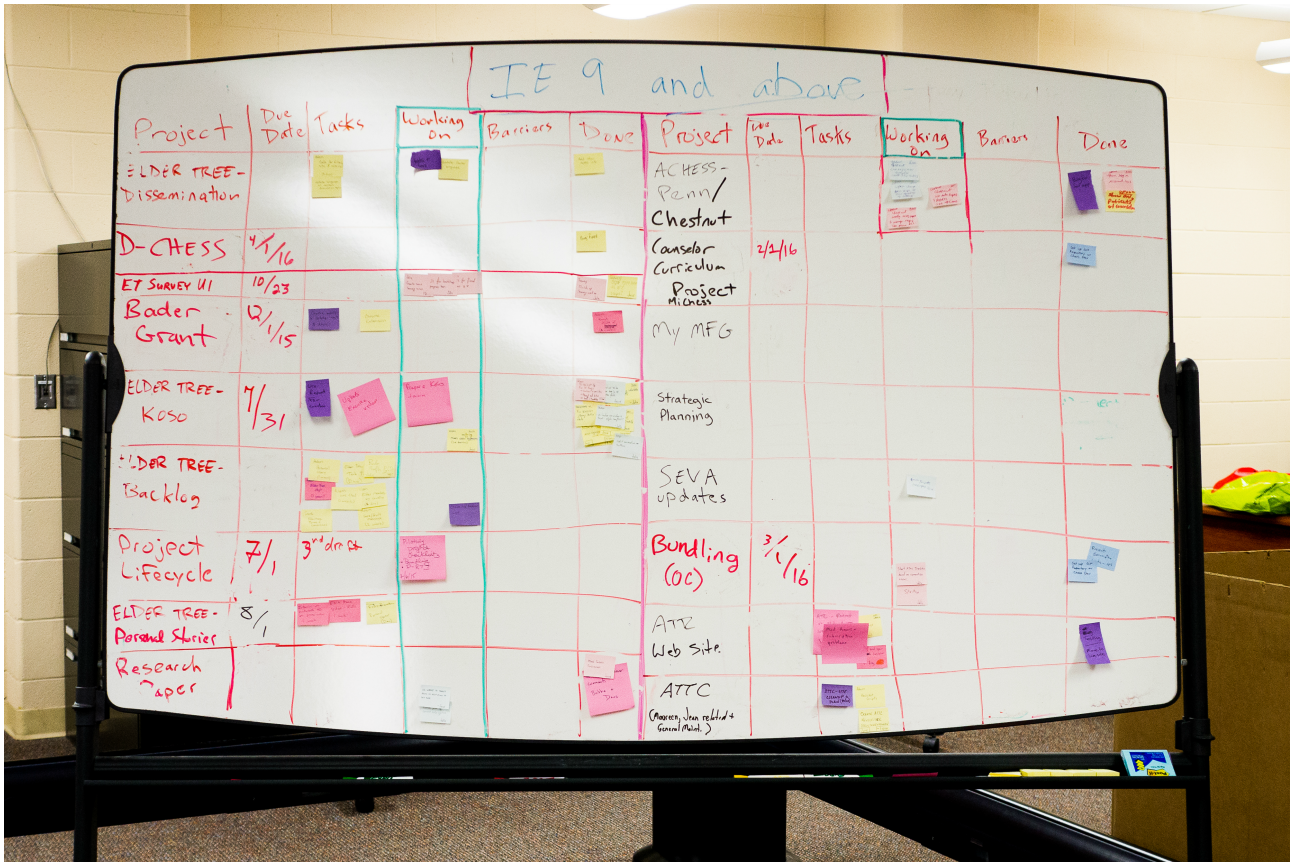
Prioritize Ideas

Our development process produced an enormous number of proposed solutions from older adults, PIs, researchers, and community partners. We had to filter these ideas so we could spend our limited resources productively. We constantly asked ourselves, "How will this feature help an older adult continue to live independently?" Remembering the overall goal of the project served as our compass. Having a strong user advocate on staff and using rapid-cycle testing also helped us filter out nonessential improvements.

To help us establish and assess priorities in the project, we used a modified agile project board (Figure 2) where we listed all the features under development with time estimates and barriers to completion. As new features were suggested, this board was an effective visual snapshot of the tech team's workload. The board was also helpful in assessing priorities with PIs.

As this project continues, we constantly reevaluate key problems. Although older adults identified the initial key problems we addressed, other sources of information have influenced us as the project has progressed. For example, it became clear that the health care industry would need to see value in the technology if it is going to pay for sustaining it. A key problem in this industry is keeping down costs. Could the technology help detect health problems that, if identified early and treated quickly, could prevent the need for costly medications and hospitalizations? We are in the later stages of developing a reporting function for health care providers and have been told by providers and insurers that it will be a very important development.

Figure 2. Photograph of the tech teams's project board.



Get Ideas From Outside the Field

We were excited about the potential of this principle to drive innovative problem solving. However, in reality we found it difficult to implement. We used this principle occasionally, such as when we used the labels of newspaper sections as models in our work, usually when we were struggling to conceptualize a feature for Elder Tree. Using this principle in its 3 phases takes an investment of time and resources, in part because it requires reading and research. We found it impossible to use this principle spontaneously in the context of a large meeting. We do see great value in this principle for future projects and plan to continue to assess the time spent on reading and researching to apply the principle against the anticipated value of the results.

You Can Never Have Enough Communication

Because of the organizational complexity of the project, we wanted to have a clear and effective communication plan in place. We feared that work on the individual research teams would proceed in silos. We adopted a communication plan to give the tech team direct, regular access to the PIs who led work on the research teams. The tech team supervisor also held regular meetings that included tech team members and change leaders to discuss development status, brainstorm ideas, and coordinate future development. The lead PI and project director made themselves available to attend these meetings when necessary to work through impediments to progress. Weekly and eventually biweekly steering committee meetings brought together the PIs from the research teams, the lead PI, the project director, and change leaders to update all attendees on progress, collect feedback, and discuss any barriers we were facing. Having many avenues of communication was a priority for us.

Ensure Rapid-Cycle Testing Is Rapid and Has Clear Goals

Of the 2 types of rapid-cycle testing we conducted (ie, usability tests and pilot tests), usability tests produced more rapid results. One-on-one usability tests are comparatively inexpensive to conduct and gave us immediate feedback on usability. These tests were usually conducted in 1 day with only 1 or 2 staff members involved. We were able to evaluate results immediately and quickly make changes and test again. For this project, we used the wireframing program Lucidchart to create paper and digital prototypes to test new concepts in addition to the fully functional website.

By contrast, the 5 pilot tests we conducted evolved into large tests and produced feedback more slowly. In the future, we would be clearer about the length of each pilot test, the features of the system being tested, and the method of collecting qualitative data at the end of each test. The first pilot test we conducted took place in 1 county for 2 weeks. At the end of the 2 weeks, we visited each participant at home and conducted a survey about how it went. The next pilot test took place in 3 counties for 1 month. Again, we visited each participant and collected survey data. Each successive test had more participants. Reaching out to each participant to survey him or her on use became a scheduling and human resource challenge. Pilot tests began to take too long, and survey data were not collected in a timely manner.

While the pilot tests provided feedback on long-term, real-life use of the website and eventually led to the creation of a beta site, they were not rapid-cycle tests. Our goals for each pilot test became less clear. Instead of testing specific elements, the pilots tested the whole system, which sometimes made it difficult to pinpoint what needed to be changed. However, our experiences with usability and pilot tests led to a clearer understanding of when and where to use specific UCD methods. For example, usability tests are a good way to test individual components of a system, whereas pilot tests are good for assessing the overall value of a system. One of our colleagues has developed a model that shows the types and sequence of technology testing within a research environment. This model will help us apply UCD methods at progressive phases of development in future projects (Table 4).

Table 4. Isham model of technology testing sequence (from feasibility to efficacy).

Feasibility [a]	Usability	Perceived usefulness	Efficacy
Does the concept show promise? Can it be built?	Can users navigate the interface?	Do users think the technology is helping?	Does the technology actually help users?
	Do they understand what is happening?	Do they want to keep using it?	
Test the concept using discussion, focus groups, and interviews with key stakeholders and end users.	Test navigation using paper prototypes, mock-ups, card sorting, and usability testing of early builds.	Longer pilot tests with users operating the system in their own environment.	Run a full experiment.

[a] The stages of technology testing generally occur in the order shown in Table 4 (ie, from feasibility to efficacy). The cost of testing generally becomes more expensive from left to right.

Conclusion

Developing the technology for this project required a constant balancing between features and simplicity. We repeatedly heard from end users that they valued simplicity over added features, while other stakeholders (community partners, PIs, National Advisory Committee members) frequently suggested adding new features. The NIATx model, with its focus on the end user, allowed us to keep the interests of older adults first and foremost and create a site that anecdotal evidence suggests does help create community and reduce isolation.

Many factors suggested that the NIATx model might be a useful framework for technology development, such as its basis in years of research about successful change projects, its origin outside the world of user-centered systems design, its simplicity, and its inclusion of a method for arriving at innovative solutions. Although we encountered challenges, we believe the NIATx model is an effective approach to UCD, especially for those not familiar with human factors or UCD principles, and we look forward to trying it again in future projects, as well as continuing to refine the use of the model throughout the development life cycle.

Acknowledgments

The Agency for Healthcare Research and Quality is the primary funder of the study (Grant No 5P50HS019917-04). Epic Systems Corporation is a secondary funder. The funders had no role in the review or approval of this manuscript. The authors also wish to thank the older adults and their informal caregivers who have not just tested but helped create Elder Tree.

Authors' Contributions

DHG Jr drafted the original manuscript. All authors read, contributed to, and approved the final manuscript.

Conflicts of Interest

The authors of the paper, except Johnson and Atwood, are members of the CHESS tech team—the group of CHESS employees who design, develop, and test CHESS ICTs. Dinauer, Isham, and Johnson have a shareholder interest in CHESS Mobile Health, a small business that develops Web-based health care technology for patients and family members. This relationship is extensively managed by the authors and the University of Wisconsin. All other authors declare that they have no competing interests.

References

1. Agency for Healthcare Research and Quality (AHRQ). Bringing Communities and Technology Together for Healthy Aging. (Grant No P50 HS1991704). 1991. URL: http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=P50+HS19917-04&QUICK=wisconsin [accessed 2015-12-12] [WebCite Cache ID 6ZBR8uT13]
2. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: A randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]

3. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: A randomized trial. Cancer 2013 May 1;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]
4. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: Results of an eHealth RCT. Health Psychol 2014 Oct;33(10):1261-1272 [FREE Full text] [doi: 10.1037/a0034216] [Medline: 24245838]
5. Gustafson D, Wise M, Bhattacharya A, Pulvermacher A, Shanovich K, Phillips B, et al. The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: A randomized controlled trial. J Med Internet Res 2012;14(4):e101 [FREE Full text] [doi: 10.2196/jmir.1964] [Medline: 22835804]
6. Gustafson DH, Hawkins R, Pingree S, McTavish F, Arora NK, Mendenhall J, et al. Effect of computer support on younger women with breast cancer. J Gen Intern Med 2001 Jul;16(7):435-445 [FREE Full text] [Medline: 11520380]
7. Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano F, et al. Impact of a patient-centered, computer-based health information/support system. Am J Prev Med 1999 Jan;16(1):1-9. [Medline: 9894548]
8. Gustafson DH, McTavish F, Gustafson DH, Mahoney JE, Johnson RA, Lee JD, et al. The effect of an information and communication technology (ICT) on older adults' quality of life: Study protocol for a randomized control trial. Trials 2015;16:191 [FREE Full text] [doi: 10.1186/s13063-015-0713-2] [Medline: 25909465]
9. Making Your Website Senior Friendly. URL: https://www.nia.nih.gov/health/publication/making-your-website-senior-friendly [accessed 2015-09-01] [WebCite Cache ID 6bE2QGe5G]
10. Vermeulen J, Neyens JC, Spreeuwenberg MD, van RE, Sipers W, Habets H, de Witte Luc P. User-centered development and testing of a monitoring system that provides feedback regarding physical functioning to elderly people. Patient Prefer Adherence 2013 Aug;7:843-854 [FREE Full text] [doi: 10.2147/PPA.S45897] [Medline: 24039407]
11. Scandurra I, Sjölinder M. Participatory design with seniors: Design of future services and iterative refinements of interactive eHealth services for old citizens. Med 2 0 2013 Dec;2(2):e12 [FREE Full text] [doi: 10.2196/med20.2729] [Medline: 25075235]
12. Zickuhr K, Madden M. Older Adults and Internet Use. Washington, DC: Pew Internet & American Life Project; 2012. URL: http://www.pewinternet.org/2012/06/06/older-adults-and-internet-use/ [accessed 2015-12-12] [WebCite Cache ID 6ZBU9UDeR]
13. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: Technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
14. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518 [FREE Full text] [Medline: 17238394]
15. Hong Y, Goldberg D, Dahlke DV, Ory MG, Cargill JS, Coughlin R, et al. Testing usability and acceptability of a web application to promote physical activity (iCanFit) among older adults. JMIR Human Factors 2014 Oct 13;1(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.3787]
16. Endsley MR. Designing for Situation Awareness: An Approach to User-Centered Design. 2nd edition. Boca Raton, FL: CRC Press; 2011.
17. Norman DA, Draper SW. User Centered System Design: New Perspective on Human-Computer Interaction. Hillsdale NJ: L. Erlbaum; 1986.
18. Iivari J, Iivari N. Varieties of user - centredness: An analysis of four systems development methods. Info Systems J 2011;21(2):125-153. [doi: 10.1111/j.1365-2575.2010.0035]
19. Vredenburg K, Mao JY, Smith PW, Carey T. A survey of user-centered design practice. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves. New York: ACM Press; 2002:471-478.
20. Nielsen J. Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. In: Bias RG, Mayhew DB, editors. Cost-Justifying Usability. San Diego, CA: Academic Press; 1994:245-272.
21. Karat J. Evolving the scope of user-centered design. Commun ACM 1997;40(7):33-38. [doi: 10.1145/256175.256181]
22. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Ã. Key principles for user-centred systems design. Behav Inf Technol 2003 Nov;22(6):397-409. [doi: 10.1080/01449290310001624329]
23. Gustafson D, Johnson K, Capoccia V, Cotter F, Ford IJ, Holloway D, et al. The NIATx Model: Process Improvement in Behavioral Health. Madison, WI: University of Wisconsin-Madison; 2011.
24. NIATx: Removing Barriers to Treatment & Recovery. URL: http://www.niatx.net/Home/Home.aspx?CategorySelected=HOME [accessed 2015-09-01] [WebCite Cache ID 6bE46m6z0]
25. Gustafson DH, Hundt AS. Findings of innovation research applied to quality management principles for health care. Health Care Manage Rev 1995;20(2):16-33. [Medline: 7607882]
26. Gawande A. Personal Best: Top Athletes and Singers Have Coaches. Should You?. 2011 Oct 3. URL: http://www.newyorker.com/magazine/2011/10/03/personal-best [accessed 2015-12-12] [WebCite Cache ID 6ZM53Z5U4]

27. Gawande A. Big Med: Restaurant Chains Have Managed to Combine Quality Control, Cost Control, and Innovation. Can Health Care?. 2012 Aug 13. URL: http://www.newyorker.com/magazine/2012/08/13/big-med [accessed 2015-06-17] [WebCite Cache ID 6ZM5NW1wP]
28. Bonkowski J. Top Sprint Cup Pit Crew Members, Teams of 2014 Named. 2014 Nov 20. URL: http://motorsportstalk.nbcsports.com/2014/11/20/top-sprint-cup-pit-crew-members-teams-of-2014-named/ [accessed 2015-06-10] [WebCite Cache ID 6ZBXe5IuG]
29. Shewhart W. Statistical Method from the Viewpoint of Quality Control. Washington, DC: Graduate School of the Department of Agriculture; 1939.
30. Deming W. The New Economics: For Industry, Government, Education. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Educational Services; 1994.
31. Patient-Centered Needs Assessment. Madison, WI: The Center for Health Enhancement Systems Studies (CHESS) URL: http://chess.wisc.edu/chess/about/needs_assessment.aspx [accessed 2015-06-10] [WebCite Cache ID 6ZBW8V1Ri]
32. Elder Tree [Home]. URL: https://eldertree.wisc.edu/Login/Login.aspx [accessed 2015-09-01] [WebCite Cache ID 6bDsZdYk5]
33. Kretzmann J, McKnight J. Building Communities from the Inside Out: A Path Toward Finding and Mobilizing a Community's Assets. Chicago, IL: ACTA Publications; 1993.
34. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]

Abbreviations

AARC: Active Aging Resource Center
ABCD: Asset-Based Community Development
CHESS: The Center for Health Enhancement Systems Studies
ICT: Information and communication technology
NASCAR: National Association for Stock Car Auto Racing
NIATx: Network for the Improvement of Addiction Treatment
PI: Principal Investigator
UCD: user-centered design

Edited by G Eysenbach; submitted 23.06.15; peer-reviewed by MY Chih, R Pak; comments to author 05.08.15; revised version received 09.09.15; accepted 05.11.15; published 14.01.16.

Please cite as:
Gustafson Jr DH, Maus A, Judkins J, Dinauer S, Isham A, Johnson R, Landucci G, Atwood AK
Using the NIATx Model to Implement User-Centered Design of Technology for Older Adults
JMIR Human Factors 2016;3(1):e2
URL: http://humanfactors.jmir.org/2016/1/e2/
doi:10.2196/humanfactors.4853
PMID:27025985

Original Paper

Changes in Default Alarm Settings and Standard In-Service are Insufficient to Improve Alarm Fatigue in an Intensive Care Unit: A Pilot Project

Azizeh Khaled Sowan[1], RN, PhD; Tiffany Michelle Gomez[2], MSc (Nursing); Albert Fajardo Tarriela[3], RN; Charles Calhoun Reed[2], RN, PhD; Bruce Michael Paper[4], BA

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

[2]Center for Excellence in Patient Care, University Health System, San Antonio, TX, United States

[3]University Health System, Transplant Cardiac Intensive Care Unit (ICU), San Antonio, TX, United States

[4]Office of Nursing Research, School of Nursing, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Corresponding Author:
Azizeh Khaled Sowan, RN, PhD
Department of Health Restoration and Care Systems Management
School of Nursing
University of Texas Health Science Center at San Antonio
Suite 2.628
7703 Floyd Curl Dr. - MC 7975
San Antonio, TX, 78229-3900
United States
Phone: 1 210 567 5799
Fax: 1 210 567 1719
Email: sowan@uthscsa.edu

Abstract

Background: Clinical alarm systems safety is a national concern, specifically in intensive care units (ICUs) where alarm rates are known to be the highest. Interventional projects that examined the effect of changing default alarm settings on overall alarm rate and on clinicians' attitudes and practices toward clinical alarms and alarm fatigue are scarce.

Objective: To examine if (1) a change in default alarm settings of the cardiac monitors and (2) in-service nursing education on cardiac monitor use in an ICU would result in reducing alarm rate and in improving nurses' attitudes and practices toward clinical alarms.

Methods: This quality improvement project took place in a 20-bed transplant/cardiac ICU with a total of 39 nurses. We implemented a unit-wide change of default alarm settings involving 17 parameters of the cardiac monitors. All nurses received an in-service education on monitor use. Alarm data were collected from the audit log of the cardiac monitors 10 weeks before and 10 weeks after the change in monitors' parameters. Nurses' attitudes and practices toward clinical alarms were measured using the Healthcare Technology Foundation National Clinical Alarms Survey, pre- and postintervention.

Results: Alarm rate was 87.86 alarms/patient day (a total of 64,500 alarms) at the preintervention period compared to 59.18 alarms/patient day (49,319 alarms) postintervention (P=.01). At baseline, Arterial Blood Pressure (ABP), Pair Premature Ventricular Contractions (PVCs), and Peripheral Capillary Oxygen Saturation (SpO2) alarms were the highest. ABP and SpO2 alarms remained among the top three at the postproject period. Out of the 39 ICU nurses, 24 (62%) provided complete pre- and postproject survey questionnaires. Compared to the preintervention survey, no remarkable changes in the postproject period were reported in nurses' attitudes. Themes in the narrative data were related to poor usability of cardiac monitors and the frequent alarms. The data showed great variation among nurses in terms of changing alarm parameters and frequency of replacing patients' electrodes. Despite the in-service, 50% (12/24) of the nurses specified their need for more training on cardiac monitors in the postproject period.

Conclusions: Changing default alarm settings and standard in-service education on cardiac monitor use are insufficient to improve alarm systems safety. Alarm management in ICUs is very complex, involving alarm management practices by clinicians, availability of unit policies and procedures, unit layout, complexity and usability of monitoring devices, and adequacy of training on system use. The complexity of the newer monitoring systems requires urgent usability testing and multidimensional interventions

to improve alarm systems safety and to attain the Joint Commission National Patient Safety Goal on alarm systems safety in critical care units.

(JMIR Human Factors 2016;3(1):e1) doi:10.2196/humanfactors.5098

KEYWORDS

cardiac monitors; default alarm settings; alarm fatigue; intensive care unit; nursing; in-service; survey

Introduction

Bedside physiologic monitors are equipped with alarm systems for patient safety and appropriate functionality. Nevertheless, the problematic high volume of false and clinically insignificant nonactionable true positive alarms—up to 99.4%—results in clinicians' failure to appropriately respond to alarms signaled from monitoring devices [1-5]. Clinicians become overwhelmed and desensitized with the number of alarms, a phenomenon known as alarm fatigue. Alarm fatigue leads to different forms of unsafe workarounds, including a delayed response, disabling alarms, turning the volume to inaudible, or adjusting alarms' settings to hazardous limits, all of which can result in missing lethal alarms. The Joint Commission (JC), which accredits and certifies health care organizations and programs in the United States, attributed alarm-related incidents and deaths to alarm fatigue and issued a 2014 National Patient Safety Goal to improve the safety of clinical alarm systems [6].

Alarm safety is a priority in intensive care units (ICUs) where alarm rates are known to be the highest [7,8]. Adjusting default alarm settings and staff education on alarm management are two strategies recommended by safety and professional organizations to reduce the number of false alarms and alarm fatigue [9-11]; however, most research on alarm safety is observational. Interventional projects that examined the effect of changing default alarm settings on overall alarm rate and on clinicians' attitudes and practices toward clinical alarms and alarm fatigue are scarce [12,13]. To address these gaps in the literature, this project aims to examine if (1) a change in default alarm settings of the cardiac monitors and (2) in-service nursing education on monitor use in an ICU would result in reducing alarm rate and improving nurses' attitudes and practices toward clinical alarms.

Methods

Design, Setting, and Sample

After obtaining Institutional Review Board (IRB) approval, this quality improvement project took place in a 20-bed transplant/cardiac ICU located at a university teaching Magnet hospital in the Southwest of the United States. The unit has 39 nurses and an average annual admission of 1500 patients. In April 2014, the unit went through three simultaneous changes, including a move to a new tower and the deployment of new bedside cardiac monitors (Philips IntelliVue MX800) and Cisco phones. The unit is an E shape and has three central nursing stations equipped with cardiac monitors (Philips IntelliVue Information Center iX) with no dedicated monitor watchers. Typical training on new medical devices includes a few hours' demonstration on the appropriate use of devices by the company representative and/or the unit nurse educators. Nurse educators also provide "if needed" support on equipment use. The interventions took place 2.5 months after the move to the new tower and the use of the new cardiac monitors.

Philips IntelliVue MX800 and the Information Center iX cardiac monitors are equipped with complex information systems with tens of main menus and as many submenus, keys, buttons, and icons to facilitate patient data surveillance and management. The monitors are operated using an interactive touch screen, a mouse and a keyboard, or a remote control. The monitors are also capable of complex functions such as lab data integration, drug calculations, guiding care through embedded clinical protocols, issuing reports and strips, presenting trended alarm data, and displaying hundreds of different alarm messages.

The Interventions

The interventions included (1) changing default settings of some parameters on the cardiac monitors and (2) re-educating transplant/cardiac ICU bedside nurses on the appropriate use of the monitors. Default settings were changed based on scientific clinical rational and recent evidence [7,12-14]. Parameters involved in the change are presented in Table 1. Change methods included the following:

1. Limit tightening.

2. Limit increase.

3. Changing the source of alarm detection to enhance alarm reading.

4. Changing the measurement mode in order to capture real conditions from different measurement sources. Measurement modes were changed from "One Source" to "Auto" (eg, System Pulse) and from "Auto" to "Enhanced" (eg, Asystole). "Auto" and "Enhanced" modes allow the monitor to look for an alternate heart rate source, such as the pulse oximeter or the arterial line if it cannot pick up a rhythm from the electrocardiogram (ECG) leads.

5. Alarm delay by increasing the period from alarm detection to announcement.

6. Disabling alarms, for example, Noninvasive Blood Pressure (NBP) Done Tone and Atrial Fibrillation (AFIB). NBP Done Tone is a nonactionable alarm announced automatically by the monitor after measuring the patient blood pressure. The AFIB alarm was disabled because it is also captured by the Irregular Heart Rate alarm. The definitions of alarm events involved in the changes are presented in Multimedia Appendix 1; some of the definitions were adapted from the IntelliVue Information Center iX Guide [15].

Changes were directed toward decreasing the number of false alarms and increasing monitoring safety. For example, although tightening the Premature Ventricular Contractions (PVCs)/minute from 10 bpm to 6 bpm is expected to increase alarm events of this parameter, this tightening was necessary for safety purposes because all other PVC-related alarms, such as Run PVCs, Pair PVCs, Bigeminy, Trigeminy, and Multiform PVCs, were disabled. Similarly, although the limit of ExtremeTachy was increased to decrease the number of false alarms, TachyClamp was tightened for safer monitoring.

7. Volume adjustment including (1) decreasing the volume of yellow alarms with moderate priority, for example, Heart Rate (HR), and (2) increasing the volume of high-priority red alarms, for example, Desaturation. Changes in alarm volume are not expected to directly affect alarm rates, but rather to focus the nurse's attention on actionable high-priority alarms for safety purposes.

All parameters involved in the change are amenable to adjustments by clinicians except for TachyClamp and ExtremeTachy, which are considered hard stops for system safety and can be adjusted only by Philips representatives (see Table 1).

The nursing unit educators conducted roaming individual in-service sessions. Educational sessions included all nurses in the unit and focused on assessment of monitor parameters, customizing parameters to be patient specific, steps of changing alarm parameters, steps of printing alarm parameters, relearning arrhythmias and changing lead analysis, and troubleshooting common alarming problems (eg, silencing alarms of monitors not connected to patients).

Procedure and Instrumentation

A team of three expert transplant/cardiac ICU nurses and a Philips representative created the list of proposed changes in parameters. This list then went through a review and approval process by all transplant/cardiac ICU physicians, nurse directors, educators, managers, and bedside nurses. The list of approved changes is presented in Table 1. After approval and before implementing any changes to bedside monitors, we invited all transplant/cardiac ICU nurses to complete a survey about nurses' attitudes and practices toward clinical alarms using an adapted version of the Healthcare Technology Foundation (HTF) National Clinical Alarms Survey [5]. A detailed description of the survey, the adaptation process, and results of the preintervention survey are presented elsewhere [16]. The postintervention survey includes three sections: (1) demographics, (2) 22 items measured on a 5-point Likert scale of agreement measuring nurses' attitudes toward clinical alarms followed by a free-text comment area, and (3) a rank section of nine items describing issues that threaten alarm recognition and response when using the cardiac monitors. The survey was followed by three additional questions related to (1) frequency of changing alarm parameters, (2) frequency of changing electrodes, and (3) adequacy of the training received on cardiac monitors.

After collecting the preintervention surveys, a Philips representative completed a unit-wide change to all monitors based on the approved list on July 1, 2014. This change was also communicated through emails, shift reports, huddles, and meetings to all transplant/cardiac ICU nurses and physicians. The unit in-service education started right after the changes in monitors' parameters and lasted for approximately two weeks. After that, an invitation to complete the postintervention survey via SurveyMonkey went out to all nurses using individual emails that included the same ID number used in the preintervention survey. Two email reminders were sent to nonrespondents to enhance the response rate.

Alarm events were measured by retrieving the audit log of the cardiac monitors from the database of the central-station monitors for 10 weeks before and 10 weeks after implementing the changes in parameters. The audit log is a chronological record of all alarm events logged by the bedside cardiac monitors.

Data Analysis

Nurse characteristics, alarm rate, and attitude toward clinical alarms were described using descriptive statistics. *Z* tests were used to measure the difference in alarm rates per patient day between the preproject and postproject periods. The change in nurses' attitudes toward clinical alarms was described using a percent change. *t* tests for paired data were used to analyze the difference in mean scores of the ranks assigned to the nine issues affecting alarm recognition (section 3 in the survey) between the preproject and postproject periods.

Table 1. Changes in default settings of the cardiac monitors at the transplant/cardiac intensive care unit.

Type of change	Parameter	Default setting	Changed to...
Limit tightening			
	PVCs[a]/minute	10 bpm	6 bpm
	TachyClamp[b]	200 bpm	180 bpm
Limit increase	ExtremeTachy[b]	20 bpm > Heart Rate High Limit	40 bpm > Heart Rate High Limit
Changing the source of alarm detection			
	ABP[c]	Source: Systolic	Source: Systolic and Mean
	NBP[d]	Source: Systolic and Mean	Source: Systolic
Changing the measurement mode			
	System Pulse[e]	SpO2[f]	Auto (from ABP, SpO2, etc)
	Asystole	Standard	Enhanced
Alarm delay	SpO2: Average[g]	No	Yes: 10 seconds
Disabling alarms			
	ST[h] Analysis[i]	On	Off
	Run PVCs	On	Off
	Pair PVCs	On	Off
	Bigeminy PVCs	On	Off
	Trigeminy PVCs	On	Off
	Multiform PVCs	On	Off
	Pause	On	Off
	Missed Beat	On	Off
	AFIB[j]	On	Off
	NBP Done Tone	On	Off
Decrease alarm volume	Yellow Alarm Volume	5	3
Increase alarm volume	Red Alarm Volume	+0	+2

[a]PVC: premature ventricular contraction.

[b]These alarms are not amenable to change by clinicians. All other alarms can be customized by clinicians based on the patient condition.

[c]ABP: arterial blood pressure.

[d]NBP: noninvasive blood pressure.

[e]If the peripheral capillary oxygen saturation (SpO2) had a poor waveform, the pulse from the pleth would not pick up and would therefore alarm. Changing to Auto allows the monitor to detect a pulse from other sources before alarming.

[f]SpO2: peripheral capillary oxygen saturation.

[g]SpO2 will be averaged over 10 seconds to determine a value instead of alarming the second the SpO2 drops below the limit. The nurse can also manually increase this to 20 or 30 seconds.

[h]ST: ST segment in the electrocardiogram.

[i]The ST Analysis alarm was disabled but should be turned on for all interventional cardiology cases (eg, require cardiac catheterization) or acute coronary syndrome patients. For these specific patients, the original limit of +/-2.0 mm should be tightened to +/- 1.6 mm as per physicians' requests.

[j]AFIB: atrial fibrillation.

Results

Nurse Characteristics

Out of the 39 transplant/cardiac ICU nurses who responded to the preintervention survey, 24 (62%) returned completed responses in the postintervention period. General characteristics of the 39 ICU nurses are described elsewhere [16]. The majority of the 24 nurse respondents were female (15/24, 63%) and worked full time (19/24, 79%). Almost half were 30-50 years old (13/24, 54%) and the other majority were less than 30 years old (10/24, 42%). Although 46% (11/24) reported having more

than 5 years of nursing experience, 79% (19/24) reported having less than or equal to 5 years of transplant/cardiac ICU experience. Chi-square tests for correlation revealed no significant differences between the 24 nurse respondents and the total 39 transplant/cardiac ICU nurses on age, gender, employment status, or total years of nursing or ICU experiences (P>.10).

Alarm Rate

Table 2 shows the number of alarms, their specific types, and difference in alarm rates per patient day for the parameters targeted in the change between the two project periods. The audit log recorded a total of 64,500 alarms at the preproject period and 49,319 at the postproject period. At baseline, Arterial Blood Pressure (ABP), Pair PVC, and Peripheral Capillary Oxygen Saturation (SpO2) alarms were the highest. ABP and SpO2 alarms remained among the top three at the postproject period. Although we disabled ten parameters (see Table 1), the data showed incomplete elimination of these alarms (see Table 2). We investigated in order to check if these alarms were activated by nurses or if they were missed from the change and discovered that one of our bedside monitors was missed from the unit change. If that monitor had been included in the change, it would have further eliminated 130 alarms (0.16 alarms/patient day) in the postproject period. These alarms included 13 Pair PVCs, 3 Multiform PVCs, 82 Missed Beat alarms, 23 Asystole alarms, and 9 ST alarms. Although we disabled the NBP Done Tone alarm, the audit log does not record this alarm. Therefore, the difference in alarm rates between the two project periods excludes the rate of that specific alarm. Using Z tests, the difference in proportions of alarm events (87.86 vs 59.18 alarms/patient day) between the two project periods was significant (P=.01), with a 24% reduction of total alarms.

Survey Results

Although all 39 nurses responded to the preintervention survey [16], we only analyzed the results of the paired sample of nurses who provided complete responses in the preproject and postproject periods (n=24).

Nurses' Attitudes and Practices Toward Clinical Alarms

The internal consistency reliability of the 22-item scale that measured attitude toward alarms using Cronbach alpha was high (.72-.75 for the pretest and post-test, respectively). Table 3 displays percentages of the 24 nurses who agreed or strongly agreed with the statements that measured attitudes toward clinical alarms and the percent change. Almost all nurses agreed/strongly agreed that nuisance alarms are frequent, disrupt patient care, and reduce trust in alarms causing caregivers to disable them (items 1, 2, and 3). Major issues threatening alarm recognition and response according to the majority of nurses in the two project periods were related to the confusion in locating the alarming device (item 4), unit layout (item 8), inadequacy of alarm systems to alert nurses of changes in patients' conditions (item 18), the lack of clinical policies and procedures on alarm management (item 21), and the inability of the newer monitoring systems to solve alarm problems (item 22). The majority of nurses were in favor of using smart alarms and central alarm management staff, and the integration of alarms to wireless devices (items 5, 6, 7, and 9).

The positive changes at the postproject period were related to the requirement to document alarm settings, the distinct outputs of medical devices, effective policies to manage alarms, and the ability of the newer systems to solve alarm problems (items 17, 20, 21, and 22). However, in this project, a positive or negative change in attitude on an item was considered clinically meaningful only if reported by at least one-third of nurses (ie, 8 nurses). Table 3 shows that the number of nurses with a change in attitude in the postproject period ranged from 0 to 6 nurses, therefore no major changes in attitude were reported.

Table 2. Difference in alarm rates between the preproject and postproject periods.

Alarm condition	Preproject period		Postproject period	
	Number of alarms	Total alarm rate/patient day	Number of alarms	Total alarm rate/patient day
ABP [a]				
Total	27,930	38.05	28,049	33.67
ABPs[b] (systolic)	13,776		14,726	
ABPm[c] (mean)	13,548		12,895	
ABP disconnect	606		428	
Pair PVCs[d,e]	8305	11.31	164	0.19
SpO2 [f]				
Total	7079	9.64	8290	9.95
SpO2	6741		7858	
SpO2r[g] (right)	338		323	
SpO2l[h] (left)	0		109	
Multiform PVCs[e]	5865	7.99	19	0.02
NBP [i]				
Total	3686	5.02	3976	4.77
NBPm[j] (mean)	1847		43	
NBPs[k] (systolic)	1837		3933	
NBPd[l] (diastolic)	2		0	
PVCs/min	3233	4.40	5330	6.39
Run PVCs high[e]	2155	2.94	23	0.03
ST[e]	1851	2.52	2609	3.13
AFIB [e,m]				
Total	1481	2.02	32	0.04
AFIB	990		26	
End AFIB[n]	491		6	
Pause[e]	1086	1.48	8	0.01
Missed Beat[e]	873	1.19	89	0.11
Asystole	323	0.44	565	0.68
Tachy [o]				
Total	292	0.39	153	0.18
Tachy	273		153	
Tachy/p[p] (tachycardia p wave)	19		0	
Vent[q] Bigeminy[e]	234	0.32	7	0
Vent Trigeminy[e]	79	0.11	0	0
Pulse	28	0.04	5	0.01
Total	64,500	87.86	49,319	59.18

[a]ABP: arterial blood pressure.

[b]ABPs: arterial blood pressure systolic.
[c]ABPm: arterial blood pressure mean.
[d]PVC: premature ventricular contraction.
[e]These are the alarms that we disabled.
[f]SpO2: peripheral capillary oxygen saturation.
[g]SpO2r: peripheral capillary oxygen saturation right.
[h]SpO2l: peripheral capillary oxygen saturation left.
[i]NBP: noninvasive blood pressure.
[j]NBPm: noninvasive blood pressure mean.
[k]NBPs: noninvasive blood pressure systolic.
[l]NBPd: noninvasive blood pressure diastolic.
[m]AFIB: atrial fibrillation.
[n]End AFIB alarm indicates the end of the AFIB status.
[o]Tachy: tachycardia.
[p]Tachy/p: tachycardia p wave.
[q]Vent: ventricular.

Narrative Data

In a previous publication, we reported detailed analysis of the narrative data provided by the 39 transplant/cardiac ICU nurses who responded to the preintervention survey [16]. Categories and themes identified in that report were related to (1) constant nuisance alarms and their effect on patient safety, (2) poor usability and complexity of medical devices, (3) the look-alike and sound-alike alarms, (4) the lack of support to the use of monitor watchers or integration of alarms into nursing call systems, and (5) unit-related factors to alarm management. The latter includes absence of policies and procedures on alarm management, the fact that unit layout may hinder response to alarms specifically when a nurse is assigned to patients who are far apart, and the need for further training on the cardiac monitors.

In the postintervention survey, 10 out of 24 (42%) nurses provided comments. These comments were matched for the preproject periods and were analyzed. Issues identified were very similar to our previous report [16] with a major focus on (1) the usability of the cardiac monitors and (2) the frequent alarms. In the postproject survey, nurses listed new cardiac monitor usability-related issues, such as the inability of the cardiac monitor to interpret ECG and nurses' inability to enter the “do not resuscitate” orders.

Importance of Alarm Issues Related to Cardiac Monitors

The respondents’ rankings of the nine statements about the importance of alarm issues specific to cardiac monitors (section 3 in the postintervention survey) is presented in Table 4. Frequent false alarms, difficulty in understanding alarm priority, and noise competition from nonclinical devices were ranked as the top three important issues interfering with alarm recognition and response in the two project periods. Difficulty in setting alarms properly because of lack of knowledge on the appropriate limits remained one of the least important issues in the postproject period. However, the lack of training on alarm systems rose from level 8 in the preintervention survey to level 4 in the postintervention survey. No significant differences were found in mean scores of the rankings between the preproject and postproject periods.

Table 3. Number and percentage of nurses who *agreed* or *strongly agreed* on the statements between the preproject and postproject periods (n=24).

Item	Statement[a]	Preproject, n (%)	Postproject, n (%)	% change[b]
1	Nuisance alarms occur frequently	24 (100)	18 (75)	-25.0
2	Nuisance alarms disrupt patient care	23 (96)	23 (96)	0
3	Nuisance alarms reduce trust in alarms and cause caregivers to inappropriately turn alarms off at times other than setup or procedural events	21 (88)	22 (92)	4.8
4	When a number of devices are used with a patient, it can be confusing to determine which device is in an alarm condition	21 (88)	19 (79)	-9.5
5	Smart alarms (eg, where multiple parameters, rate of change of parameters, and signal quality are automatically assessed in their entirety) would be effective to use for improving clinical response to important patient alarms	20 (83)	17 (71)	-15.0
6	Central alarm management staff responsible for receiving alarm messages and alerting appropriate staff is helpful	19 (79)	18 (75)	-5.3
7	Smart alarms (eg, where multiple parameters, rate of change of parameters, and signal quality are automatically assessed in their entirety) would be effective to use for reducing false alarms	19 (79)	16 (67)	-15.8
8[c]	Unit layout does interfere with alarm recognition and management	18 (75)	18 (75)	0
9	Alarm integration and communication systems via pagers, cell phones, and other wireless devices are useful for improving alarms management and response	15 (63)	17 (71)	13.3
10[c]	Nearly all alarms are actionable (requiring the nurse to respond and take an action)	14 (58)	14 (58)	0
11	Alarm sounds and/or visual displays of the current monitoring systems and devices should clearly differentiate the priority of alarm	13 (54)	14 (58)	7.7
12	Properly setting alarm parameters and alerts is overly complex in existing devices	13 (54)	13 (54)	0
13	Clinical staff is sensitive to alarms and responds quickly	13 (54)	15 (63)	15.4
14[c]	When a lethal alarm sounds, it is clearly and quickly recognized and immediate action is taken to address the alarm	12 (50)	14 (58)	16.7
15	Environmental background noise has interfered with alarm recognition	12 (50)	15 (63)	25.0
16	Alarm sounds and/or visual displays should be distinct based on the parameter or source (eg, device)	12 (50)	16 (67)	33.3
17[d]	There is a requirement in my unit to document that the alarms are set and are appropriate for each patient	11 (46)	18 (75)	63.6
18[d]	The alarms used on my unit are adequate to alert staff of potential or actual changes in a patient's condition	10 (42)	9 (38)	-10.0
19	There have been frequent instances where alarms could not be heard and were missed	8 (33)	8 (33)	0
20[d]	The medical devices used on my unit all have distinct outputs (ie, sounds, repetition rates, visual displays) that allow users to identify the source of the alarm	8 (33)	15 (63)	87.5
21[d]	Clinical policies and procedures regarding alarm management are effectively used in my unit	6 (25)	11 (46)	83.3
22	Newer monitoring systems (eg, < 3 years old) have solved most of the previous problems we experienced with clinical alarms	1 (4)	6 (25)	500

[a]Edited and used with permission from the Healthcare Technology Foundation (HTF) 2011.

[b]Percent change = ((y2 - y1) / y1) × 100.

[c]These are the new statements that we added to our survey. They do not exist in the original HTF survey.

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

Table 4. Importance of alarm issues related to the cardiac monitors (n=24).

Item	Statement	Preproject		Postproject		
		Item response, mean	Mean ranking[a]	Item response, mean	Mean ranking	*P*
1[b]	Frequent false alarms, which lead to reduced attention or response to alarms when they occur	2.40	1	3.40	1	.11
2[b]	Difficulty in understanding the priority of an alarm	3.00	2	4.32	2	.07
3[b]	Noise competition from nonclinical alarms and pages	3.95	3	4.55	3	.50
4[c]	Lack of available policy on appropriate alarm parameters for individualized patients	4.40	4	5.80	9	.08
5[c]	The need to frequently reset alarm settings every time they revert back to default when the monitor is disconnected from the patient	4.42	5	5.16	5	.24
6[d]	Difficulty in hearing alarms when they occur, especially from outside patient room	4.47	6	5.37	7	.36
7[d]	Difficulty in setting alarms properly because of the complexity of the monitor	4.84	7	5.21	6	.70
8[b]	Lack of training on alarm systems	4.90	8	4.70	4	.83
9[d]	Difficulty in setting alarms properly because of lack of knowledge on the appropriate limits for my patient condition	5.42	9	5.58	8	.75

[a]Item response means were ranked from 1 (most important) to 9 (least important).

[b]These statements were adopted from the Healthcare Technology Foundation (HTF) survey.

[c]These statements were added to the survey to reflect the cardiac monitors.

[d]These statements were modified from the HTF survey. Original statements were as follows: item 6 “Difficulty in hearing alarms when they occur”; items 7 and 9 “Difficulty in setting alarms properly.”

Practices Related to Clinical Alarms and Training on Cardiac Monitors

The data showed great variation among nurses in terms of changing alarm parameters (see Figure 1). More than one-third of nurses reported not adjusting alarm parameters in the preproject period. Despite the in-service, 25% (6/24) of nurses sustained the same practice in the postproject period. Additionally, only 38% (9/24) of nurses individualized parameters based on the patient’s vital signs in the two project periods.

The frequency of replacing patients' electrodes also varied. However, only 54% (13/24) of nurses changed them daily during the two project periods (see Figure 2).

In regard to the training needed on the cardiac monitors, the majority of nurses indicated that they did not receive sufficient training on the central and bedside monitors (19/24, 79% and 16/24, 67%, respectively) in the preproject period. Despite the in-service, almost half of the nurses specified their need for more training in the postproject period (see Figure 3).

Figure 1. Percentage of nurses who modify the bedside alarm parameters in the pre- and postproject periods (n=24).

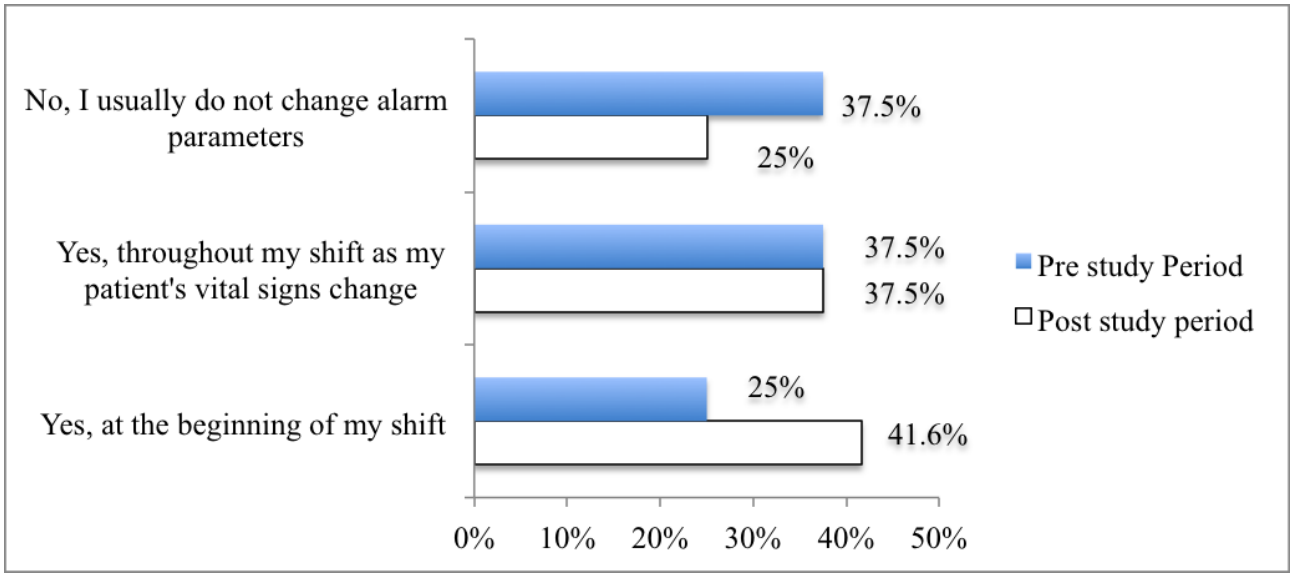


Figure 2. Percentage of nurses who replace patients' electrodes in the pre- and postproject periods (n=24).

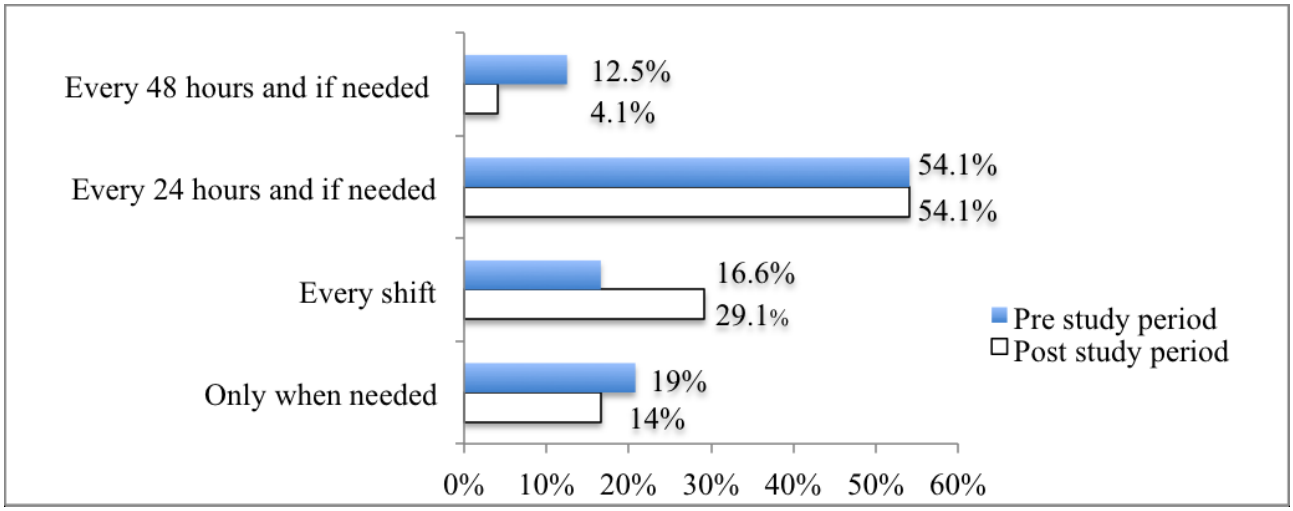
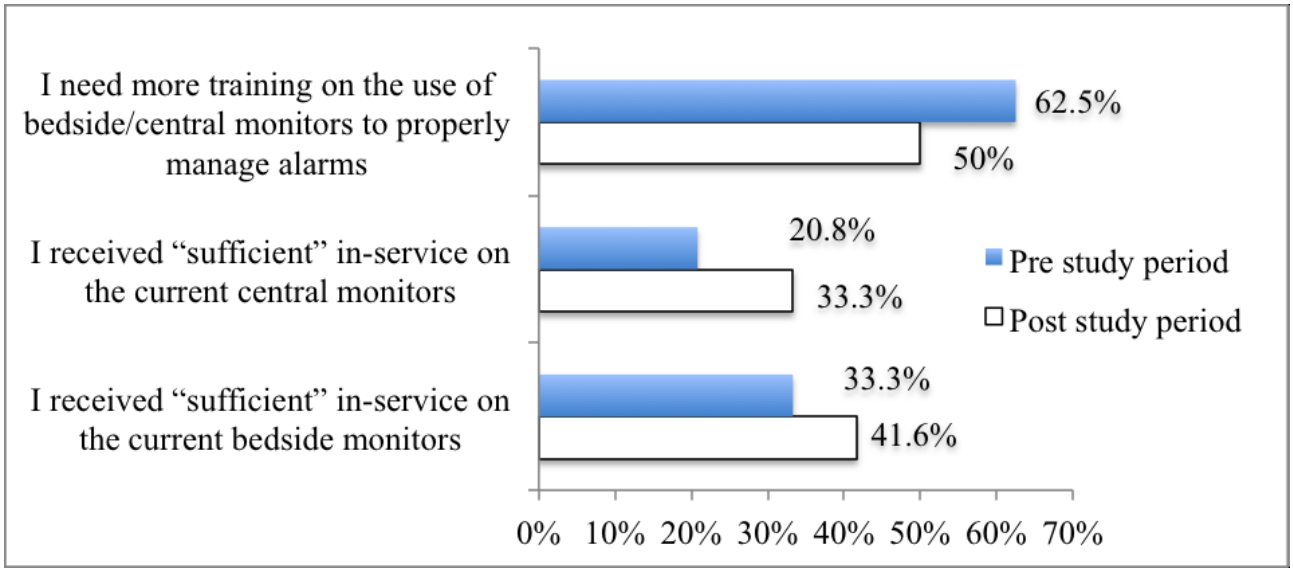


Figure 3. Percentage of nurses who received and needed monitor training in the pre- and postproject periods (n=24).



Discussion

Overview

Examining the effect of interventions targeting alarm systems safety on nurses' attitudes toward alarms and alarm fatigue related-practices is critical to evaluate improvements in the safety of these systems. Our unit-wide changes in default alarm settings of cardiac monitors significantly reduced 24% of the total number of the target alarms. However, changing default alarm settings, the subsequent reduction in alarm rate, and the in-service education on alarm management were insufficient to improve nurses' attitudes toward alarms, alarm fatigue, or maintaining best clinical practices.

Principal Findings and Future Directions

Finding alarms for parameters that were turned off supports the fact that bedside nurses customize patient alarms. ABP and SpO2 alarms were among the highest in the preproject and postproject data. The specific types of alarms (eg, Arterial Blood Pressure Mean [ABPm] and Arterial Blood Pressure Systolic [ABPs] alarms; see Table 2) can guide future initiatives on further alarm reduction. Future studies may examine if all specific types of alarms were necessary to be monitored for the patient. This may reveal alarm overuse and explain the high number of alarms. For example, clinicians need to determine if there is a need to monitor SpO2 right (SpO2r) and SpO2 left (SpO2l) for every patient.

The 65% increase in PVCs/min alarms (from 3233 to 5330) is expected because we tightened the parameter. However, the 41% increase in ST alarms (from 1852 to 2609) was unexpected given that we disabled this parameter. Changing this parameter to "On" by nurses is a plausible interpretation for such an increase. The ST parameter includes 12 leads. It would be helpful to analyze if nurses turned on the ST alarm as per the recommended cases by physicians and according to the suggested limits, which leads they adjusted, or if they overused the alarms. Correlating alarm rates and conditions to reliable monitoring conditions is critical and has not yet been investigated. For example, ST monitoring is not recommended in cases when arrhythmias such as atrial flutter and fibrillation are present or if the patient is continuously ventricularly paced.

These cases will result in frequent false nonactionable ST alarms.

The unexpected increase in Asystole alarms (from 323 to 565, 75%) can be related to acuity of patients' conditions and infrequent electrode placement. Another possible explanation from our observation is not adjusting the Pace Mode to "On" in the monitor for patients with temporary pacemakers who keep alarming Asystole. Our results also showed that nurses do not follow the unit protocol (ie, every 24 hours and if needed) when changing leads; enforcement of this policy should take place [11]. In a telemetry unit, proper skin preparation and electrode placement resulted in a significant reduction of ECG alarms [17].

Despite the significant reduction in alarm rate, key issues causing alarm fatigue and reducing trust in alarm systems according to nurses were the high frequency of nuisance alarms, the confusion in locating the alarming device, a unit layout that hinders alarm response, the inadequacy of alarm systems to alert nurses of changes in patients' conditions, the lack of clinical policies and procedures on alarm management, and the complexity of the newer monitoring systems. These multiple issues emphasize the fact that alarm management is very complex in ICUs. On the other hand, and similar to our previous results [16], the narrative data attributed nurses' frustration and desensitization to alarms to poor usability of the cardiac monitoring systems.

It seems that the complexity of these monitors require interactive, well-designed, and periodic training. Our in-service, though individualized and focused on changing and individualizing alarm parameters and troubleshooting common problems, was insufficient to enhance appropriate monitor use. This is supported by finding that 50% of nurses believed they still needed training on cardiac monitor use and suggests (1) the need for usability testing of cardiac monitors, (2) the use of super-users, and (3) a competency checklist that includes key features for monitor use. Usability studies may reveal the complexity of the monitors, lack of knowledge about some features, or inappropriate use of the monitors. Studies supported the lack of clinicians' awareness about, and understanding of, the complexity of cardiac monitors [16,18]. On the other hand, the wide variations in nurses' practices and lack of adherence to protocols related to frequency of changing patients' electrodes and parameters are major factors behind frequent nuisance alarms. Best practices should be enforced through unit policies. Inconsistent practices are indicative of the need for further education on appropriate programming and use of monitoring devices.

Summary

Cardiac monitors are receiving increased attention in ICUs because of the high number of alarms triggered by these devices compared to other alarm-equipped ICU devices (ie, infusion pumps, dialysis pumps, and mechanical ventilators) [19,20]. Unlike other studies [12,13], our multimethod approach in addressing alarm fatigue was unsuccessful in improving attitudes toward alarms and safety practices. This can be related to the difference in patient population, the type and complexity of cardiac monitors in use, and nurses' noncompliance to best practices related to a lack of unit policies on alarm management. Inconsistent practices related to alarm management by medical, surgical, and ICU nurses have been reported [16,21]. Studies also support the perceived relationships between inappropriate setting of alarm parameters and the high number of false alarms in ICUs [22]. On the other hand, unlike many other observation-based clinical alarms safety studies [23,24], we measured alarm events using an objective data source of the audit log. The audit log provides a comprehensive record of all cardiac monitor alarms, except for the NBP Done Tone.

If we included the alarms from the monitor missed from the unit change and the NBP Done Tone alarms, the actual alarms' reduction rate would be more than 24%. The inconsistency in applying the same unit of analysis in measuring alarm rates hinders further comparison across alarm safety studies.

Limitations

The sample of nurses was small. This limited examining the statistical difference in attitudes toward clinical alarms. Although we achieved a significant reduction in alarm rate, we did not correlate that to the acuity of patient conditions preintervention and postintervention. Our description of alarms was limited to the alarms that we targeted for change. The audit log of the cardiac monitors records all types of physiologic alarms and all technical alarms. Analyzing other alarms may provide more insight into the total number of alarms triggered by the cardiac monitors. Our study was limited to alarms from the cardiac monitors and did not include other frequently used alarming devices in ICUs, such as infusion pumps or ventilators. However, cardiac monitors were the devices associated with the highest number of death cases in the US Food and Drug Administration data [19].

Conclusions

Changing default alarm settings and standard in-service education on cardiac monitor use are insufficient to improve alarm systems safety. Alarm management in ICUs is very complex, involving alarm management practices by clinicians, availability of unit policies and procedures, unit layout, complexity and usability of monitoring devices, and adequacy of training on systems use. The complexity of the newer monitoring systems requires urgent usability testing. Multidimensional interventions are needed to improve alarm systems safety and attain the Joint Commission National Patient Safety Goal on alarm systems safety in critical care units.

Acknowledgments

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

Conflicts of Interest

None declared.

Multimedia Appendix 1

Definitions of alarm events for parameters involved in the change.

[PDF File (Adobe PDF File), 32KB - humanfactors_v3i1e1_app1.pdf]

References

1. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg 2006 May;102(5):1525-1537. [doi: 10.1213/01.ane.0000204385.01983.61] [Medline: 16632837]
2. Clark T, David Y. Impact of Clinical Alarms on Patient Safety. Plymouth Meeting, PA: ACCE Healthcare Technology Foundation; 2006. URL: http://thehtf.org/white%20paper.pdf [accessed 2015-09-02] [WebCite Cache ID 6bFAJpkaI]
3. Phillips J, Barnsteiner JH. Clinical alarms: Improving efficiency and effectiveness. Crit Care Nurs Q 2005;28(4):317-323. [Medline: 16239820]
4. Atzema C, Schull MJ, Borgundvaag B, Slaughter GR, Lee CK. ALARMED: Adverse events in low-risk patients with chest pain receiving continuous electrocardiographic monitoring in the emergency department. A pilot study. Am J Emerg Med 2006 Jan;24(1):62-67. [doi: 10.1016/j.ajem.2005.05.015] [Medline: 16338512]
5. Funk M, Clark JT, Bauld TJ, Ott JC, Coss P. Attitudes and practices related to clinical alarms. Am J Crit Care 2014 May;23(3):e9-e18. [doi: 10.4037/ajcc2014315] [Medline: 24786820]
6. The Joint Commission: Sentinel Event Alert. Oak Brook, IL: The Joint Commission; 2013 Apr 08. Medical device alarm safety in hospitals URL: http://www.jointcommission.org/assets/1/18/SEA_50_alarms_4_5_13_FINAL1.PDF [accessed 2015-01-02] [WebCite Cache ID 6VHNuN0Aa]
7. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg 2009 May;108(5):1546-1552. [doi: 10.1213/ane.0b013e31819bdfbb] [Medline: 19372334]
8. Siebig S, Kuhls S, Imhoff M, Gather U, Schölmerich J, Wrede CE. Intensive care unit alarms--How many do we need? Crit Care Med 2010 Feb;38(2):451-456. [doi: 10.1097/CCM.0b013e3181cb0888] [Medline: 20016379]
9. Joint Commission on Accreditation of Healthcare Organizations. The Joint Commission announces 2014 National Patient Safety Goal. Jt Comm Perspect 2013 Jul;33(7):1,3-1,4. [Medline: 24137865]
10. A Siren Call to Action: Priority Issues From the Medical Device Alarms Summit. Arlington, VA: Association for the Advancement of Medical Instrumentation; 2011. URL: http://www.aami.org/htsi/alarms/pdfs/2011_Alarms_Summit_publication.pdf [accessed 2015-01-02] [WebCite Cache ID 6VHNSNksF]
11. American Association of Critical-Care Nurses: Practice Alert. Aliso Viejo, CA: American Association of Critical-Care Nurses; 2013. Alarm management URL: http://www.aacn.org/wd/practice/docs/practicealerts/alarm-management-practice-alert.pdf [accessed 2015-05-10] [WebCite Cache ID 6VHOOtZLJ]
12. Graham KC, Cvach M. Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010 Jan;19(1):28-34; quiz 35 [FREE Full text] [doi: 10.4037/ajcc2010651] [Medline: 20045845]
13. Dandoy CE, Davies SM, Flesch L, Hayward M, Koons C, Coleman K, et al. A team-based approach to reducing cardiac monitor alarms. Pediatrics 2014 Dec;134(6):e1686-e1694 [FREE Full text] [doi: 10.1542/peds.2014-1162] [Medline: 25384493]
14. Welch J. An evidence-based approach to reduce nuisance alarms and alarm fatigue. Biomed Instrum Technol 2011;Suppl:46-52. [doi: 10.2345/0899-8205-45.s1.46] [Medline: 21599481]
15. IntelliVue Information Center iX: Instructions for Use. Release A. 01. 1st edition. Andover, MA: Philips Medical Systems; 2012.
16. Sowan AK, Tarriela AF, Gomez TM, Reed CC, Rapp KM. Nurses' perceptions and practices toward clinical alarms in a transplant cardiac intensive care unit: Exploring key issues leading to alarm fatigue. JMIR Hum Factors 2015 Mar 16;2(1):e3. [doi: 10.2196/humanfactors.4196]
17. Walsh-Irwin C, Jurgens CY. Proper skin preparation and electrode placement decreases alarms on a telemetry unit. Dimens Crit Care Nurs 2015;34(3):134-139. [doi: 10.1097/DCC.0000000000000108] [Medline: 25840128]
18. Korniewicz DM, Clark T, David Y. A national online survey on the effectiveness of clinical alarms. Am J Crit Care 2008 Jan;17(1):36-41 [FREE Full text] [Medline: 18158387]
19. FDA Patient Safety News: Show #106. Silver Spring, MD: US Food and Drug Administration; 2011 Jan. Alarming monitor problems URL: http://www.fda.gov/downloads/Safety/FDAPatientSafetyNews/UCM417915.pdf [accessed 2015-09-02] [WebCite Cache ID 6bFBLdWA5]
20. Bridi AC, Louro TQ, da Silva RC. Clinical alarms in intensive care: Implications of alarm fatigue for the safety of patients. Rev Lat Am Enfermagem 2014;22(6):1034-1040 [FREE Full text] [doi: 10.1590/0104-1169.3488.2513] [Medline: 25591100]
21. Gazarian PK, Carrier N, Cohen R, Schram H, Shiromani S. A description of nurses' decision-making in managing electrocardiographic monitor alarms. J Clin Nurs 2015 Jan;24(1-2):151-159. [doi: 10.1111/jocn.12625] [Medline: 24813940]

22. Varpio L, Kuziemsky C, MacDonald C, King WJ. The helpful or hindering effects of in-hospital patient monitor alarms on nurses: A qualitative analysis. Comput Inform Nurs 2012 Apr;30(4):210-217. [doi: 10.1097/NCN.0b013e31823eb581] [Medline: 22156767]
23. Albert NM, Murray T, Bena JF, Slifcak E, Roach JD, Spence J, et al. Differences in alarm events between disposable and reusable electrocardiography lead wires. Am J Crit Care 2015 Jan;24(1):67-73; quiz 74. [doi: 10.4037/ajcc2015663] [Medline: 25554556]
24. Gazarian PK. Nurses' response to frequency and types of electrocardiography alarms in a non-critical care setting: A descriptive study. Int J Nurs Stud 2014 Feb;51(2):190-197. [doi: 10.1016/j.ijnurstu.2013.05.014] [Medline: 23810495]

Abbreviations

ABP: arterial blood pressure
ABPm: arterial blood pressure mean
ABPs: arterial blood pressure systolic
AFIB: atrial fibrillation
ECG: electrocardiogram
HR: heart rate
HTF: Healthcare Technology Foundation
ICU: intensive care unit
IRB: Institutional Review Board
JC: Joint Commission
NBP: noninvasive blood pressure
NBPd: noninvasive blood pressure diastolic
NBPm: noninvasive blood pressure mean
NBPs: noninvasive blood pressure systolic
PVC: premature ventricular contraction
SpO2: peripheral capillary oxygen saturation
SpO2l: peripheral capillary oxygen saturation left
SpO2r: peripheral capillary oxygen saturation right
ST: ST segment in the electrocardiogram
Tachy: tachycardia
Tachy/p: tachycardia p wave
Vent: ventricular

Edited by G Eysenbach; submitted 02.09.15; peer-reviewed by P Gazarian; comments to author 07.10.15; revised version received 15.10.15; accepted 16.10.15; published 11.01.16.

Please cite as:
Sowan AK, Gomez TM, Tarriela AF, Reed CC, Paper BM
Changes in Default Alarm Settings and Standard In-Service are Insufficient to Improve Alarm Fatigue in an Intensive Care Unit: A Pilot Project
JMIR Human Factors 2016;3(1):e1
URL: http://humanfactors.jmir.org/2016/1/e1/
doi:10.2196/humanfactors.5098
PMID:27036170

Original Paper

In the Loop: The Organization of Team-Based Communication in a Patient-Centered Clinical Collaboration System

Allison M Kurahashi[1], MEd; Peter B Weinstein[2], BEng, MHSc; Trevor Jamieson[3,4,5], MD; Jennifer N Stinson[6,7], RN-EC, PhD, CPNP; Joseph A Cafazzo[2,8,9], PhD, PEng; Bhadra Lokuge[1], MHS; Plinio P Morita[2], MSc, PhD, PEng; Eyal Cohen[6,8,10,11], MSc, MD; Adam Rapoport[11,12,13,14], MHSc, MD; Andrea Bezjak[15], MSc, MD; Amna Husain[1,14], MPH, MD

[1]Temmy Latner Centre for Palliative Care, Mount Sinai Hospital, Toronto, ON, Canada
[2]Healthcare Human Factors, Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[3]Division of General Internal Medicine, St. Michael's Hospital, Toronto, ON, Canada
[4]Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada
[5]Department of Medicine, University of Toronto, Toronto, ON, Canada
[6]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada
[7]Lawrence S Bloomberg, Faculty of Nursing, University of Toronto, Toronto, ON, Canada
[8]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[9]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada
[10]Commonwealth Fund/Canadian Foundation for Healthcare Improvement, Commonwealth Fund, New York, NY, United States
[11]Department of Paediatrics, University of Toronto, Toronto, ON, Canada
[12]Paediatric Advanced Care Team, The Hospital for Sick Children, Toronto, ON, Canada
[13]Emily's House Children's Hospice, Toronto, ON, Canada
[14]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[15]Department of Radiation Oncology, Princess Margaret Cancer Centre, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Amna Husain, MPH, MD
Temmy Latner Centre for Palliative Care
Mount Sinai Hospital
4th Floor, Box 13
60 Murray Street
Toronto, ON,
Canada
Phone: 1 416 586 4800 ext 7884
Fax: 1 416 586 4804
Email: amna.husain@utoronto.ca

Abstract

Background: We describe the development and evaluation of a secure Web-based system for the purpose of collaborative care called Loop. Loop assembles the team of care with the patient as an integral member of the team in a secure space.

Objective: The objectives of this paper are to present the iterative design of the separate views for health care providers (HCPs) within each patient's secure space and examine patients', caregivers', and HCPs' perspectives on this separate view for HCP-only communication.

Methods: The overall research program includes cycles of ethnography, prototyping, usability testing, and pilot testing. This paper describes the usability testing phase that directly informed development. A descriptive qualitative approach was used to analyze participant perspectives that emerged during usability testing.

Results: During usability testing, we sampled 89 participants from three user groups: 23 patients, 19 caregivers, and 47 HCPs. Almost all perspectives from the three user groups supported the need for an HCP-only communication view. In an earlier prototype, the visual presentation caused confusion among HCPs when reading and composing messages about whether a message was visible to the patient. Usability testing guided us to design a more deliberate distinction between posting in the Patient and

Team view and the Health Care Provider Only view at the time of composing a message, which once posted is distinguished by an icon.

Conclusions: The team made a decision to incorporate an HCP-only communication view based on findings during earlier phases of work. During usability testing we tested the separate communication views, and all groups supported this partition. We spent considerable effort designing the partition; however, preliminary findings from the next phase of evaluation, pilot testing, show that the Patient and Team communication is predominantly being used. This demonstrates the importance of a subsequent phase of the clinical trial of Loop to validate the concept and design.

(JMIR Human Factors 2016;3(1):e12) doi:10.2196/humanfactors.4996

KEYWORDS

collaborative care; patient-centered care; patient engagement; chronic disease; communication; Internet communication tools; Internet communication technologies

Introduction

Overview

As the complexity of health care increases, we are recognizing the limits of current models of program-centered and specialty-centered care [1-3]. Patient-centered care and patient engagement have the potential to substantially improve outcomes in the health care system [4-6]. The penetration of Internet and mobile technologies makes it possible to envision new systems for interactive communication that follow the patient across the continuum of care. In this paper, we present aspects of the design, development, and evaluation of such a system. The system, called Loop, uses social networking principles to assemble the patient's actual team of care and include the patient as an integral member of the team for the purpose of collaborative care.

The Gap

In the United States, 84% of all health care spending in 2006 was for the 50% of the population who have one or more chronic medical conditions [7,8]. In Canada, chronic disease contributes disproportionately to the total economic cost of illness [9,10]. Globally, chronic disease is predicted to increase both in prevalence and complexity. "The most common chronic condition experienced by adults is multimorbidity, the coexistence of multiple chronic diseases or conditions" [11]. These are patients with complex chronic disease who require multiple health providers and have unique needs, disabilities, or functional limitations [12]. Currently, health care is organized in organizational and disease-specific silos that the patient moves across frequently and unpredictably, eliciting a broad call for transformative solutions [13,14]. Wagner's chronic care model [15-17], endorsed in several countries including the United States and Canada, proposes a roadmap for effective management that calls for "planned, proactive seamless care in which the clients are full participants in managing their care and are supported to do this at all points by the system" [18]. However, there are few systems to enable engagement and collaboration. Understanding the gap and potential solution grew from our team's experience, which spans diverse populations with chronic and complex care needs including home palliative care, cancer care, acute to ambulatory care transitions, adolescents and young adults with cancer (AYAC), and children with medical complexity (CMC). Lack of communication is a problem identified across all these populations; fostering communication is a key process if we are to achieve continuity of care and comanagement [19,20]—a goal endorsed by all stakeholders [21,22]. Comanagement, or collaborative care, requires more than a passive sharing of electronic health records (EHRs). It requires ongoing, interactive, and contextual communication among team members [1]. A report from the American Medical Informatics Association's 2013 Policy Meeting on patient-centered care highlights this: "EHRs are necessary but not sufficient to engage patients and foster improvements in the quality of care . . . health information needs to flow across the health care continuum" [6].

The Solution: Loop

The evidence supports collaborative care as the keystone to chronic disease management, the patient and caregiver as integral partners in care, and communication as central to achieving these objectives [4,13,15]. We propose a solution using emerging social networking technologies: a Web-based clinical collaboration system for complex chronic disease patients. In Loop, each team of care centered on a patient, or Patient Loop, consists of the patient, the caregiver, and the health care providers (HCPs) involved in the patient's care. Each Patient Loop is a secure space partitioned from every other Patient Loop, and those with access must be involved in the patient's care and authenticated to join. While users are not provided with specific instructions, Loop is designed to encourage them to communicate questions, updates, and clarifications about care plans. Loop allows team members, including the patient, to indicate their preferences and check their understanding of the care plan. The communication is visible to the team but specific members of the team may be tagged. Therefore, in terms of types of communication there could be an exchange between patient (or caregiver) and HCP or between HCPs in the team. The purpose of Loop is for team members to arrive at care plans together and work toward a shared set of goals.

Rationale for a Web-Based Clinical Collaboration System

Several studies report the desirability, acceptability, and manageability of messaging systems focused on patient-physician communication [23-25]. Two studies evaluating patient-HCP messaging systems did not find

detectable differences in the volume of communication with these systems [23,24]. Results show participants had increased satisfaction with communication, improvement in workflow, and overall positive attitudes towards online communication [23,24,26].

We conducted a search for existing systems primarily in the United States and Canada and approached vendors and groups working in the communication space [27-30]. In the context of large health care organizations in the United States, communication is embedded within each organization's information technology. However, this does not work outside the confines of a large health network in the United States or in the contexts of other countries. For example in Canada, patients move across the single-payer system without the restrictions imposed by insurance providers or organizations. We studied a number of messaging tools developed for use within the confines of a hospital in a large urban center in Canada [2]. Separate development has led to multiple tools for similar functions with no reach beyond the organization. HCPs reported that tools are not integrated into their work flow, resulting in decreased efficiency and tools being used in incorrect ways [31]. We embarked on the research and development of Loop when communication tools for direct patient care were just emerging. Our literature review revealed a few communication tools with potential to extend beyond organizational boundaries; however, they are limited in a number of ways [27-30]. A patient-held record (PHR) may have potential to be used as a communication tool, but existing PHRs are institutionally sponsored, limited to the institution's patients, and do not have the functionality of providing a space for HCPs to communicate for collaborative care [32,33]. EHRs focus on transmitting medical reports as a means of communication without enabling interactive exchange. Still other tools limit communication to certain groups (eg, HCPs only) or to certain forms (eg, private one-on-one messaging) [26]. None of the tools had the integrated functionality we envisioned: a focus on communication for direct patient care in the community, a networking structure, and separate but integrated communication spaces for patients and HCPs. Recently, a handful of tools that overlap some of the functions of Loop have emerged [34-37]. Our program of research on Loop, with its iterative user-driven development and its robust evaluation, contributes important learning to this emerging field of eHealth tools for care coordination.

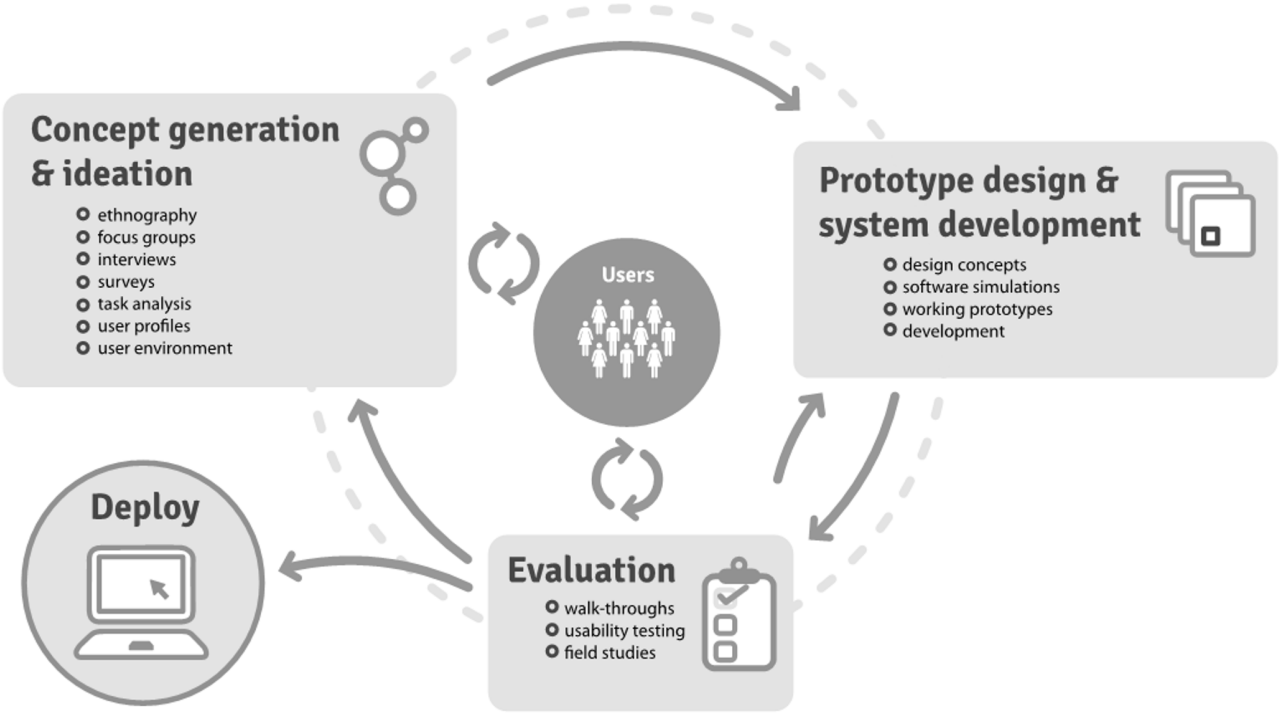
Existing systems continue to be organization-centric and propagate a model that is minimally collaborative, excludes key players, and is ill-suited to the complexity of health care [31,38]. Loop is interactive, asynchronous communication that enables the patient's team of care to be assembled, no matter what their profession, where they practice, or what their organizational affiliation is; and it includes patients, caregivers, and health care professionals in the communication. The Web-based platform allows Loop to reach beyond organizational boundaries. In future phases, we will use plug-in or application programming interface technology to link to the different EMRs across organizations. We envision that Loop will serve as a communication layer linked to other eHealth tools in a personalized dashboard.

The Development of Loop

In line with existing recommendations to rigorously evaluate eHealth systems throughout all stages of their life cycles [24], we chose a sequential plan of research following the Medical Research Council framework for complex interventions [39]. We embedded an iterative stakeholder engagement process based on user-centered design (UCD) [40,41] and participatory design methods (Figure 1). User or end-user refers to patients, caregivers, and HCPs who would use Loop in planning or coordinating care. Participatory design calls for the engagement of clinicians, researchers, developers, designers, end-users, and the technology itself throughout development [42]. Thus research and development have been integrally linked, and the various research activities have been continuous and reflexive. Through this research spanning more than 5 years, we have developed and tested the Loop prototype in simulated and real settings.

The concept of open versus private communication within the team has been at the core of the development of Loop. At inception, the research team had an idea that open communication between the members of the team of care, regardless of what their role or where they practiced, would be transformative. Previous literature has indicated that having clinical discussions in the presence of patients and families during bedside rounds improves communication and transparency [43,44]. Despite these benefits, the authors report parent and HCP concern about negative emotional responses and confusion that may result from technical discussion [43,44] and the need for "pre-rounding" or "re-rounding" away from patients and families to have uninhibited conversations [44]. Prior to the usability testing phase that is the focus of this paper, all user groups endorsed two separate communication spaces within a patient's secure space: one for the entire team including patients and caregivers and another for HCPs only. We carried this knowledge forward when creating prototypes by including an option for HCP-only communication. The perspectives on the two separate views that emerged in usability testing are the focus of this paper.

Figure 1. User-centered design process extracted from McCurdie et al [27].



Objectives

The objectives of this paper are to present the iterative design of the separate view for HCPs within each patient's secure space and examine patient, caregiver, and HCP perspectives on this separate view. While providing feedback about the visual design of the separate view during usability testing, participants also shared perspectives about HCPs communicating with each other without patients or caregivers able to view the communication within Loop. The focus of the analysis presented in this paper is derived solely from the usability testing, and we limit our description to the usability testing phase.

Methods

Summary of Phases of Work

Our search for existing systems found none with team-based communication that included both the patient and the team of HCPs, was cross-organizational, and followed the patient across the entire health system. We used UCD methods (Figure 1) to engage the final users of the product as active participants in the design process and gather user needs as product requirements [41]. Specifically, we employed the following components of UCD: (1) ethnography [45], (2) affinity diagramming [46], (3) cooperative prototyping [47], (4) dramatic simulation activities [47], (5) usability testing and prototyping [48], and (6) pilot testing. This paper focuses on usability testing while briefly discussing the other activities for context.

Usability testing followed a descriptive qualitative method [49] and content analysis, which aims to summarize the informational content of verbal and visual data [50,51]. This analysis was reflexive and interactive throughout usability testing.

Population

We recruited a convenience sample of participants from the following populations: adult cancer, adolescents and young adults with cancer (AYAC), and children with medical complexity (CMC). We recruited patients with cancer, caregivers of patients with cancer, and HCPs representing a variety of disciplines involved in cancer care. In the CMC area, we recruited parents of CMC patients and HCPs involved in their care. We obtained relevant institutional review board approvals and informed consent from all participants.

Usability Testing and Prototyping

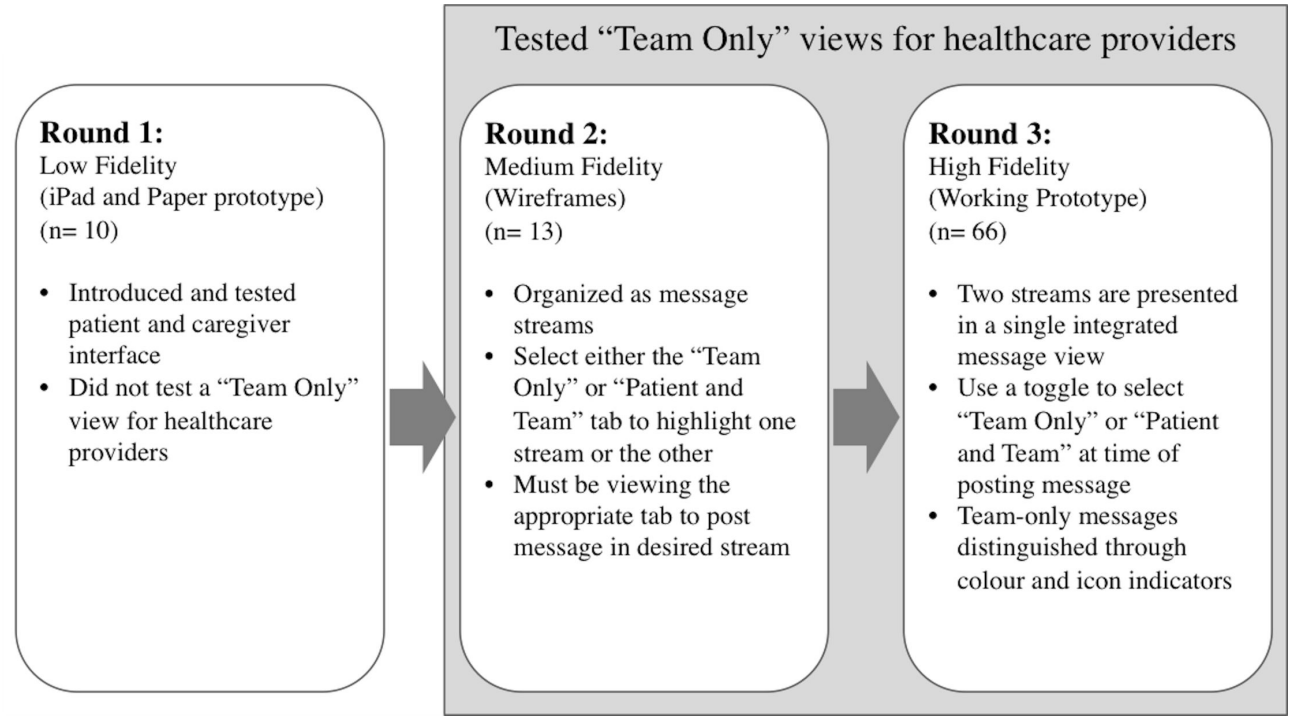
Usability Testing Procedure

During three rounds of usability testing in simulation labs, participants were asked to follow the think-aloud protocol while interacting with prototypes of progressing fidelity [48]. The prototypes were prepopulated with messages based on realistic patient stories and served as the foundation for participants to interact with and respond to. A facilitator provided task-oriented scenarios guiding participant interactions with the system and asked questions about participant experiences. Data were collected using screen and audio capture and by note-takers in an adjacent observation room. In addition, we tested the prototype offsite following the same simulation protocol with a different sample of patients and caregivers in their homes and HCPs in their practice settings. Offsite usability testing occurred concurrent with and in between rounds of simulation testing in labs. In all instances, participants were asked to complete a pretest survey for information on demographics and comfort with technology. Each session of usability testing involved a unique participant with the exception of one caregiver who participated in two usability testing sessions but is counted as one participant. All interviews were transcribed verbatim.

Applying Usability Testing Feedback to Prototype Development

A basic interactive prototype was created using Axure RP version 6.5. Prototyping early gave participants something to respond to when providing feedback about major design principles and required features. This set of specifications using screenshots and detailed descriptions informed the design and development of low-, medium-, and high-fidelity prototypes within cycles of usability testing as described above. A low-fidelity prototype focused on the introduction of a patient and caregiver interface and was a necessary step in the evolution to the later prototypes. The low-fidelity prototype did not evaluate separate streams in the HCP view. The interactive medium- and high-fidelity prototypes were the first instances where the usability and acceptability of HCP-only messaging could be tested (Figure 2).

Figure 2. Usability testing and prototype progression.



Qualitative Analysis of Usability Testing

The content of usability testing interviews was presented and discussed in weekly project team meetings. Consensus was achieved on design principles that were tracked in a user specification document that served as an audit trail for the process. Additionally, emerging concepts and major decisions of the team were captured in meeting notes. Interview transcripts were independently coded in NVivo version 10 (QSR International) by three reviewers, who met initially to arrive at consensus for a coding framework and continually to discuss subsequent coding application. At key points, two senior team members (AH and JS) reviewed the framework. The ongoing process of review has given the team a grounding in the data to inform further in-depth analysis focused on emerging concepts. Usability testing interview data were reviewed and selectively coded to identify participant perceptions about private messaging among HCPs and open communication with all team members including the patient and caregiver. Quotes were extracted from references coded in the preliminary categories of *visibility of messages*, *team composition*, *composing a message*, and *medical terminology*. As themes emerged, queries were run with the keywords *private*, *conversation*, *confusion*, and *anxiety* to identify any additional quotes related to open and closed communication. Through this process, two reviewers refined the initial categories into emergent themes.

Results

Population

Across all the activities, we had a convenience sample of 150 participants from the CMC, AYAC, and adult cancer populations (tables 1-4). A subset composed of 89 participants took part in usability testing. In this subset, there were 23 patients, 19 caregivers, and 47 HCPs. Results of the usability testing and its impact on development of the prototype are described together because one activity continually informed the other. With regard to access to technology, Internet penetration at home ranged from 91% to 100% across populations and user groups in this sample. Although the numbers are small in each user subgroup, the findings suggest trends: HCPs had the most use of computers and access to Internet both at work and home, and AYAC patients had the most comfort with smartphones and social media.

Table 1. Number of participants involved across all activities.

Data collection activity	Population	Role		Activity total
		Health care provider	Patient/ caregiver	
Focus groups, interviews, and ethnography				
	Adult cancer	14	9	
	AYAC	7	0	
	CMC	5	0	
	Total	26	9	35
Usability testing [a]				
	Adult cancer	19	20	
	AYAC	16	15	
	CMC	12	7	
	Total	47	42	89
Pilot testing				
	Adult cancer	6	3	
	AYAC	6	3	
	CMC	6	2	
	Total	18	8	26
Total		91	59	150

[a]Findings of the usability testing are the focus of this paper.

Table 2. Patient participant profile data (usability testing only).

		Adult cancer N=8	CMC N=0	AYAC N=15
Female, n (%)		5 (62)	—	5 (33)
Age, years, median (range)		61 (40-79)	—	17 (15-26)
Diagnosis, n (%)				
	Breast cancer	0 (0)	—	—
	Colorectal cancer	0 (0)	—	—
	Lung cancer	2 (25)	—	—
	Ovarian cancer	1 (12)	—	—
	ALL	—	—	3 (20)
	AML	—	—	2 (13)
	Ewing sarcoma	—	—	1 (7)
	Rhabdomyosarcoma	—	—	1 (7)
	Non-Hodgkin lymphoma	—	—	1 (7)
	Osteosarcoma	—	—	2 (13)
	Other	5 (62)	—	5 (33)
Use a computer at work/school, n (%)		5 (63)	—	13 (87)
Use a computer at home, n (%)		7 (88)	—	14 (93)
Use the Internet at home, n (%)		7 (88)	—	14 (93)
Comfortable using, n (%)			—	
	Computer	5 (63)	—	15 (100)
	Smartphone	3 (38)	—	15 (100)
	Internet	6 (75)	—	15 (100)
	Email	7 (88)	—	14 (93)
	Instant messaging	4 (50)	—	15 (100)
	Social media	2 (25)	—	13 (87)
Hours spent on computer per day				
	<1	2 (25)	—	0 (0)
	1-7	5 (63)	—	12 (80)
	>7	1 (13)	—	3 (20)
Hours spent on Internet per day				
	<1	2 (25)	—	1 (7)
	1-7	6 (75)	—	12 (80)
	>7	0 (0)	—	2 (13)

Table 3. Caregiver participant profile data (usability testing only).

		Adult cancer N=12	CMC N=7	AYAC N=0
Female, n (%)		7 (58)	6 (86)	—
Age, years, median (range)		56.5 (31-72)	37 (32-45)	—
Caregiver type, n (%)				
	Spouse	4 (33)	0 (0)	—
	Son/daughter	5 (42)	0 (0)	—
	Mother/father	1 (8)	7 (100)	—
	Other	2 (17)	0 (0)	—
Use a computer at work/school, n (%)		10 (91)	6 (86)	—
Use a computer at home, n (%)		11 (92)	7 (100)	—
Use the Internet at home, n (%)		11 (92)	6 (100)	—
Comfortable using, n (%)				—
	Computer	11 (92)	7 (100)	—
	Smartphone	10 (83)	5 (71)	—
	Internet	11 (92)	7 (100)	—
	Email	11 (92)	7 (100)	—
	Instant messaging	11 (92)	7 (100)	—
	Social media	6 (50)	4 (57)	—
Hours spent on computer per day				—
	<1	1 (8)	0 (0)	—
	1-7	8 (67)	3 (44)	—
	>7	3 (25)	4 (57)	—
Hours spent on Internet per day				—
	<1	1 (8)	0 (0)	—
	1-7	10 (83)	6 (86)	—
	>7	1 (8)	1 (14)	—

Table 4. Health care provider participant profile data (usability testing only).

		Adult cancer N=19[a]	CMC N=11[a]	AYAC N=16[a]
Female, n (%)		13 (68)	10 (91)	14 (87)
Age, n (%)				
	20-29	0 (0)	1 (9)	1 (6)
	30-39	6 (32)	2 (18)	5 (31)
	40-49	6 (32)	4 (36)	7 (44)
	50-59	3 (16)	3 (27)	2 (12)
	60-69	4 (21)	1 (9)	1 (6)
Years in health care, median (range)		15 (3-40)	20 (3-35)	18.5 (2.5-39)
Profession, n (%)				
	Family physician	7 (37)	0 (0)	—
	Community nurse	0 (0)	2 (18)	—
	Palliative care physician specialist	2 (10)	0 (0)	—
	Medical oncologist	1 (5)	0 (0)	—
	Other specialist	6 (32)	1 (9)	—
	Case manager	1 (5)	3 (27)	—
	Other	2 (10)	3 (27)	—
	General pediatrician	0 (0)	2 (18)	—
	Physician	—	—	5 (31)
	Advanced practice nurse	—	—	8 (50)
	Nurse	—	—	2 (12)
	Psychologist	—	—	1 (6)
Use a computer at work/school, n (%)		19 (100)	11 (100)	—
Use a computer at home, n (%)		19 (100)	11 (100)	—
Use the Internet at home, n (%)		19 (100)	11 (100)	—
Comfortable using, n (%)				
	Computer	19 (100)	11 (100)	—
	Smartphone	18 (95)	9 (82)	—
	Internet	19 (100)	11 (100)	—
	Email	18 (100)	10 (100)	—
	Instant messaging	17 (94)	10 (100)	—
	Social media	8 (44)	5 (50)	—
Hours spent on computer per day				
	<1	0 (0)	0 (0)	—
	1-7	13 (68)	6 (55)	—
	>7	6 (32)	5 (46)	—
Hours spent on Internet per day				
	<1	0 (0)	1 (9)	—
	1-7	14 (74)	7 (64)	—
	>7	5 (26)	3 (27)	—

[a]Percentages are calculated based on the number of answers submitted. Not all questions were completed by all participants.

Objective 1: Usability Testing and Prototyping

The medium-fidelity prototype was the first iteration to have interfaces for patients and caregivers in addition to the interface for HCPs (Figure 2). In the prototype (Figure 3), the HCP view was organized as two streams of messages: one with messages visible to the patient and the other to HCPs only. Each stream was given a different visual treatment, and HCPs were able to select the stream of conversation to join from this view. The patient and caregiver view had only one stream of messages. This organization caused confusion for some HCPs, who found it hard to tell whether the patient was involved in a conversation. Therefore, this design did not meet our objective of an intuitive user experience. Further analysis indicated that for HCPs the distinction was more important while composing and sending messages than while viewing messages.

In the high-fidelity prototype (Figure 4), we incorporated this learning by removing the visual treatment of the two streams and introducing a prominent toggle (*Patient and Team* and *Team Only*) in the compose message box prompting HCPs to make a selection at the time of posting the message. Once posted, any message for HCPs only is distinguished by an icon. The reply message is by default an HCP Team Only message unless Patient and Team is actively selected in the toggle.

Figure 3. Medium-fidelity prototype of health care provider view with two message streams distinguished by visual treatment. Scenario and mock-up based on actual patient case.

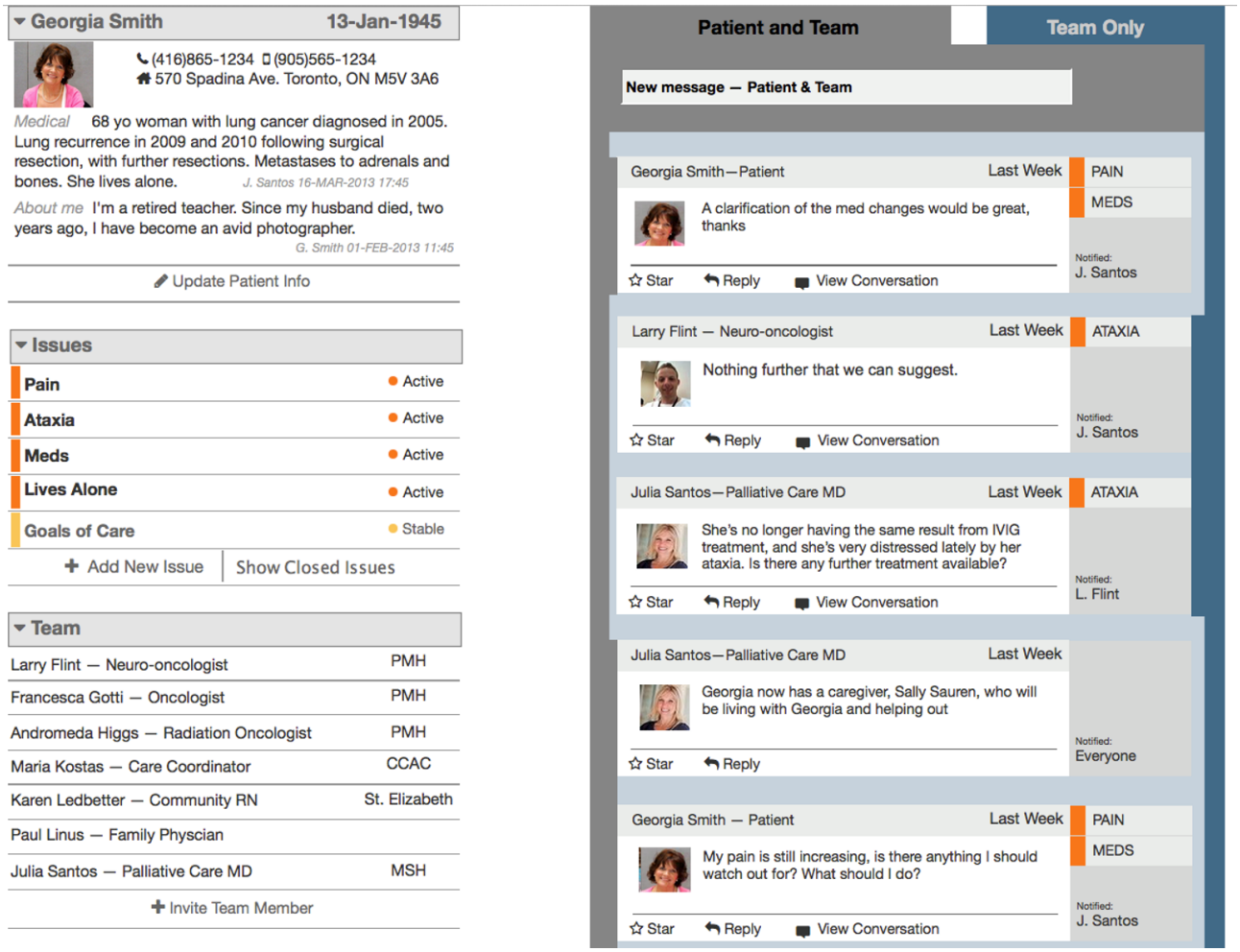
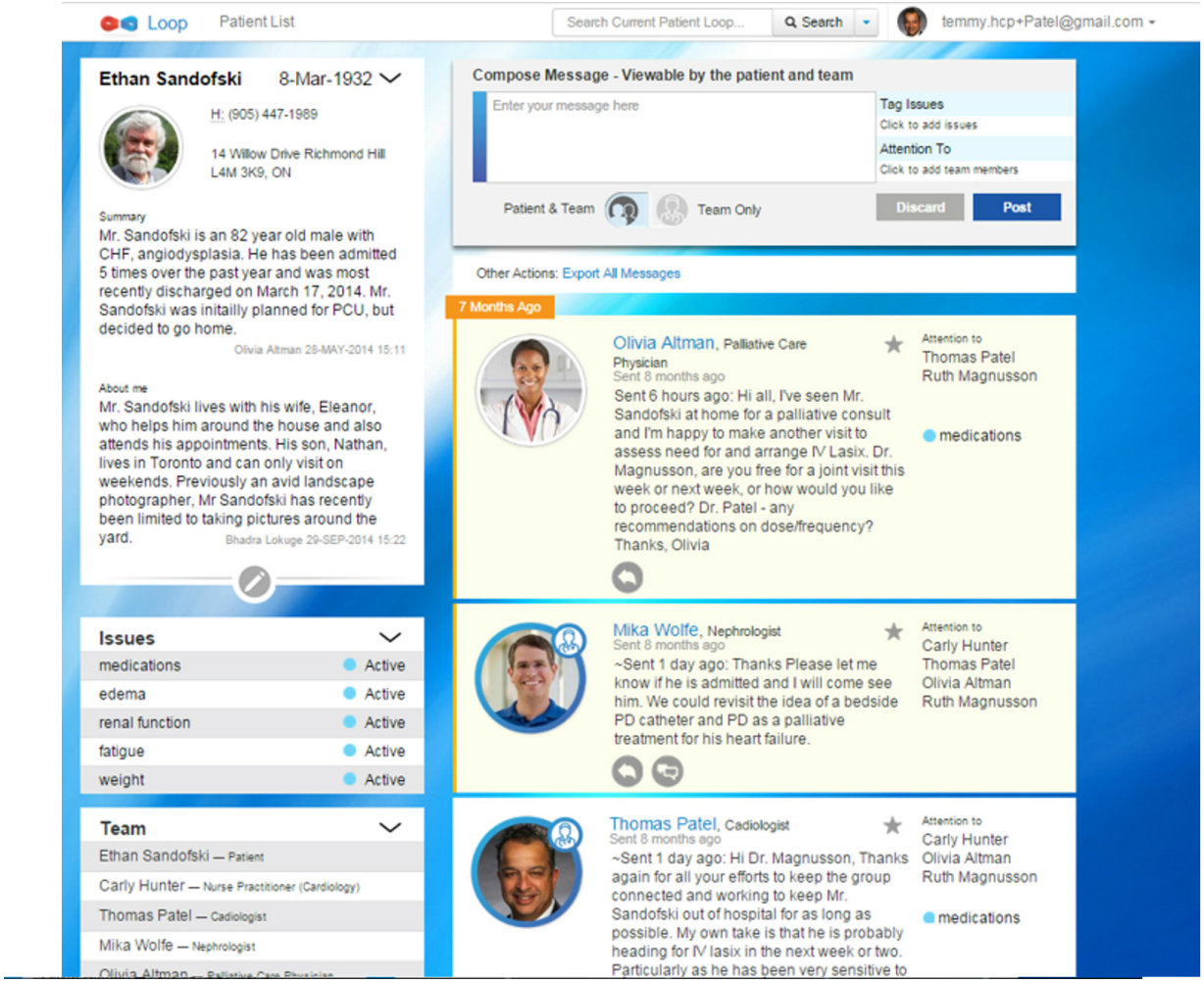


Figure 4. High-fidelity prototype version of Loop with Patient and Team and Team Only toggle from message compose box.



Objective 2: Qualitative Analysis of Usability Testing

Analysis of usability testing transcripts found that the concept of open messaging between the team and patient and caregiver was new to participants across all user groups. In both the medium- and high-fidelity prototype cycles of usability testing, the vast majority expressed the need for a separate space for HCP communication within the Patient Loop, with only a few participants concerned that this would disenfranchise patients and caregivers.

Those who expressed a need for a separate space for HCP communication had two main reasons. First, HCPs may communicate more freely and efficiently if patients are not part of the conversation.

> *In a Team Only circumstance, you can probably say things in a little bit more free-form or with less restraint. And that's partly because you need to be that frank. You need to say listen, this is very worrisome, don't know what to say to mom, let's have a conversation about this and this is where that Team Only option is good.* [HCP #1, CMC]
>
> *Theoretically, I buy into the idea that the patient should be part of everything. But it really does change what you're able to put in the message. So, in the real world, when you're really in a hurry, and you really want to get some stuff out there, you might not be able to.* [HCP #2, adult cancer]
>
> *And we're not necessarily accustomed to talking to other health professionals with the patient aware of every word that's said or the way it's said or the way it could easily be misinterpreted . . . those kinds of things.* [HCP #3, adult cancer]

Second, messages could contain information that causes confusion or anxiety for patients and caregivers. For example, patients and caregivers may experience confusion after viewing a message about a preliminary stage of planning.

> *If it's something that is in very preliminary discussion and it's not a possibility but it's a thought, I probably would not wanna be privy to that.* [Patient #1, AYAC]
>
> *. . it would be beneficial for the doctors to talk amongst themselves before they give you an answer that might mislead you to [think] something else . . .* [Patient #2, AYAC]
>
> *You don't necessarily want to hear everything that the doctors are discussing . . . You want to hear the end discussion, you don't want to be more confused.* [Caregiver #1, CMC]
>
> *In circumstances where you want to have an internal conversation about consultation and you're not really*

> *sure what the plan is and it's not necessarily for the family to know because the family typically wants to know what the plan is. They don't necessarily want to be part of the particulars around the plan . . .* [HCP #1, CMC]

In addition, patients and caregivers may experience anxiety if messages contain new information about their disease or treatment.

> *Well, this [the message stream in Loop] is very detailed about what happened and how it went and I think if a patient is willing to read all this, is comfortable reading that, it is good. But what if a patient is not comfortable reading whatever details there are and how bad it is?* [Caregiver #2, adult cancer]

> *. . . It's not saying that the patient is not going to get it, but there's a way to disclose that information where the patient could have some family there and instead of doing it this way, where they might just get it as an email. It's so much more impersonal than actually sitting down and having a conversation.* [HCP #4, AYAC]

> *So, if it's only a medical [term] that she's unlikely to understand, or might get freaked out about.* [HCP #5, adult cancer]

> *. . . there are some times where you're in a formative stage and it's not probably to the patient's interest to talk about really what's his prognosis and have we determined that, before I say it to him.* [HCP #6, adult cancer]

These perspectives supporting a separate communication space for HCPs based on protection of the patient and efficiency for the HCP should be considered in light of the few divergent perspectives favoring all communication be open to patients and caregivers.

> *I understand both sides being health care people want to talk about health care issues. And if you're talking about how to disclose a diagnosis to a patient, you can't have the patient reading that. At the same time, I think that [separate views] makes this a health care provider–favored tool rather than a patient-favored tool.* [HCP #7, CMC]

> *. . so the only thing I'm questioning with this is whether it should go to the patient, but . . . I'm assuming the point of it all is, that's why I'm just trying to . . . I was just sort of struggling with how to phrase it. [Pause] . . . because it's such a sensitive issue but then . . . I mean I think it should because it's obviously a team issue that the patient's brought forward. And I assume that part of this is to be completely transparent. Is that to have all discussions and that for the patient to know . . . what team members are saying . . . that there's a transparency to it.* [HCP #8, AYAC]

Patients, caregivers, and HCPs all believe that there are different considerations, conventions, and language governing conversations between HCPs versus those between HCPs and patients. The challenge for Loop is how to best weave these conversations in one platform, respecting the prevailing perspectives but keeping true to the aim of changing the status quo as it relates to patient engagement.

Discussion

Principal Findings

This paper describes the perspectives of patients, caregivers, and HCPs on an HCP-only communication space within a patient's secure communication space. All user groups, including patients and caregivers, support HCP-only communication. Usability testing informed the design of this partition in successive prototypes. The overall program of research has explored this core concept as well. At inception, the team had an idea that open communication, where all messages were visible to the patient and the entire professional team, would democratize communication and mitigate the hierarchies that exist in health care. Consistent with prior ethnographic work, usability testing showed that almost all end-users including patients and caregivers endorsed a separate HCP message view. The challenge in Loop is that it must serve the communication needs of all its user groups: patients, caregivers, and HCPs. Loop must be able to accommodate the different considerations that govern communication between HCPs versus communication between patients and HCPs. At the same time, it must provide the flexibility to engage patients in their plan of care. We found similar perspectives reflected in the literature. While patients want to be engaged in health care decisions, they trust their clinicians to have the knowledge and skills to arrive at and propose appropriate options for care [52]. Patients want to be engaged in the decision-making process according to their preferences for receiving communication [53]. In Loop, this preference can be accommodated within each team. Each Patient Loop is expected to be a self-regulating microenvironment based on the characteristics, context, and behavior of team members. It is expected that team agreements and roles will change and evolve over time through interactions in Loop.

Although the data from pilot testing in real-world teams will be the focus of a subsequent paper, preliminary results suggest that the majority of messages exchanged are between patients and HCPs and therefore are in the open Patient and Team view. This illustrates the need for phases of evaluation of implementation and effectiveness that we are currently conducting. Our sequential approach to evaluation is supported by two reviews of health information technology used to facilitate communication; both call for evaluation that uses methodological standards such as the Medical Research Council framework [54,55]. Additionally, a 2015 scoping review of information and communication technology (ICT) supports our UCD approach [54]; only 6% of 350 studies identified for inclusion evaluated usability of the tool to any degree. The authors of the review point out the need for usability testing: "*This is disturbing since usability is an important factor for the acceptability of ICT by its users, and the lack of attention paid to usability in the reviewed studies indicates that there would be much to be gained from this*" [54].

Limitations

One limitation of expert and user feedback is that end-users may not know what they need until using a system in practice [56]. Additionally, end-users often have wish lists that are specific to their context, which make contradictory demands on the system and make it less usable for other end-users [57]. Therefore, we analyzed and prioritized the feedback in a group with clinical, research, development, and design representation. Usability testing in a lab or simulated setting does not allow for evaluation of how a system would be used in a real-world setting and how it fits into workflow. This will be tested in subsequent phases of our research.

The sample may be biased toward individuals who were more engaged, favorably disposed to technology, and functionally capable. We will need to address the challenge of accessibility, adoption, and scaling in the next phase of the work.

Comparison With Prior Work

We found no studies examining the issue of separate message views for patients and HCPs in a team-based communication system. In addition to the studies on patient-physician communication referenced previously [52,53], there is a body of literature that examines the perceived benefits and concerns associated with bedside rounds conducted in the presence of the patient and caregiver [58]. Grzyb et al [43] surveyed parents of children admitted to the neonatal intensive care unit and medical trainees who rotated through the unit to solicit views on parents being in attendance at rounds. Stickney et al [44] interviewed parents of children admitted to an intensive care unit and HCPs (nurse, residents, fellows, and attending intensive care unit physicians) about parents' and providers' goals and expectations for participation in morning rounds. Our findings echo the findings in these papers: HCPs did not like discussing unfavorable prognoses in front of parents and felt that discussion among providers was inhibited. They worried about information being misinterpreted and a "negative emotional response to unwelcome news" and felt that parents attending rounds made for longer rounds [44]. Parents had polarized views on whether they should be given bad news during rounds, felt it might be upsetting to hear health care providers express uncertainty about their child's condition or treatment [43], and were concerned about being confused by the technical nature of the discussion [44]. Parents felt more included in their child's care when they were present for bedside rounds [44].

There are only a handful of systems like Loop (ie, tools for cross-sectoral collaboration with team-based communication as their focus) in use today. The application of networking technologies to communication in health care is an emerging field. As Bates states in a recent article: "*If organizations want to succeed in improving quality and reducing costs, providing better care coordination is one of the most important keys. However, the electronic health records of today do not yet truly enable care coordination. Even the leading US organizations in care coordination do not yet have robust electronic tools for doing this—making this a key frontier for clinical informatics*" [38].

Any intervention developed for the purpose of clinical communication about the patient and involving the patient must be patient-centered. The 2015 scoping review of ICT states that "*hardly any of the interventions could be regarded as 'fully' person-centered care (PCC) meeting the 3 routines of initiating the partnership (patient narratives), working the partnership (shared decision making), and safeguarding the partnership (documenting the narrative)*" [54]. Loop facilitates each of these processes. Its authors further state that "*shared decision making, personal information sharing, and setting up a care plan enabled by ICT seem to be relatively new*" [54].

In describing a system like Loop, it is important to address the question of feasibility. We acknowledge that the problem of poor communication is not just a technology problem. Implementing Loop requires considering the characteristics of individuals, organizations, incentives, and policies. At the individual level, HCPs fear a tsunami of electronic messages or an erosion of the rules that have traditionally governed patient and HCP communication. Evidence from prior studies and our pilot testing of Loop itself does not show the overall volume of messages to be increased with the introduction of electronic communication [23,24]. However, there is no denying that Loop challenges the system to rethink the role of the patient and how HCPs communicate.

On the organizational and health system level, the accountability and payment incentive systems are often based on organizationally defined objectives. Coordination is not adequately compensated, posing an existential challenge to Loop. However, regardless of incentives, many HCPs spend a significant portion of their time chasing information and connecting with people to deliver safe, quality care. If Loop can save time in doing these tasks, the impact is obvious.

On a policy level, scaling Loop must consider complexities related to ownership of the system, privacy, data sharing, and regulatory approval. In a system organized in silos of funding, the method of payment for a cross-organizational tool like Loop is unclear. A broad coalition of partners is necessary for a collaborative project but difficult to translate into an effective governance and payment model.

Conclusions

The development process of Loop shows the importance of grounding eHealth systems in clinical practice and patient experiences. Only through a robust research and UCD process is it possible to identify underlying issues and constraints. The core concept of open versus private communication evolved from the initial vision for open communication to partitioning the space to create an HCP-only view based on user perspectives and the preliminary pilot testing showing that open communication is predominantly being used. This demonstrates that the next phase of a clinical trial of Loop is a critical step in validation of the UCD. In the trial, we will evaluate whether the functionalities that emerged through our approach so far translate as intended in clinical practice and patient experience of Loop.

Acknowledgments

The authors would like to acknowledge the support of Christopher Obwanga, Fatima Kanji, Renee Desjardins, Margaret van Wyk, and Russell Goldman. This work was supported by the Ramona Rull Karson Innovation Grant of the Canadian Cancer Society (grant #701156), the Canadian Institutes of Health Research Pfizer Operating Grant in Disease Prevention and Management Competition (FRN 122030), and the Canadian Institutes of Health Research eHealth Innovation Catalyst Grant (CEH 126529). Eyal Cohen is supported by The Commonwealth Fund, a private independent foundation based in New York, NY. The views presented here are those of the author and not necessarily those of The Commonwealth Fund, its directors, officers, or staff.

Authors' Contributions

All authors made substantial contributions to conception and design or acquisition, analysis, and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

References

1. Stille CJ, Primack WA. Interspecialty communication: old problem, new hope? Arch Intern Med 2011 Jul 25;171(14):1300. [doi: 10.1001/archinternmed.2011.326] [Medline: 21788553]
2. Popovici I. Patient-centered perspectives of communication and handover between the emergency department and general internal medicine. Toronto, ON: University of Toronto; 2011. URL: https://tspace.library.utoronto.ca/handle/1807/31390 [accessed 2015-07-10] [WebCite Cache ID 6ZvI7N9U8]
3. Smith M, Saunder R, Stuckhardt L, McGinnis JM, editors. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington DC: National Academies Press; 2013. URL: http://www.nap.edu/read/13444/chapter/1 [WebCite Cache ID 6ZvIpZa9a]
4. Craig C, Eby D, Whittington J. Care coordination model: better care at lower cost for people with multiple health and social needs. Cambridge, MA: Institute for Healthcare Improvement; 2011. URL: http://www.ihi.org/resources/Pages/IHIWhitePapers/IHICareCoordinationModelWhitePaper.aspx [accessed 2015-07-10] [WebCite Cache ID 6ZvJAfgnS]
5. Laurance J, Henderson S, Howitt PJ, Matar M, Al Kuwari H, Edgman-Levitan S, et al. Patient engagement: four case studies that highlight the potential for improved health outcomes and reduced costs. Health Aff 2014 Sep;33(9):1627-1634. [doi: 10.1377/hlthaff.2014.0375] [Medline: 25201668]
6. Brennan PF, Valdez R, Alexander G, Arora S, Bernstam EV, Edmunds M, et al. Patient-centered care, collaboration, communication, and coordination: a report from AMIA's 2013 Policy Meeting. J Am Med Inform Assoc 2015 Apr;22(e1):e2-e6. [doi: 10.1136/amiajnl-2014-003176] [Medline: 25359545]
7. Anderson G. Chronic care: making the case for ongoing care. Princeton, NJ: Robert Wood Johnson Foundation; 2010. URL: http://www.rwjf.org/content/dam/farm/reports/reports/2010/rwjf54583 [accessed 2015-07-10] [WebCite Cache ID 6ZvJMn2HB]
8. Chronic disease prevention and health promotion: chronic disease overview. Atlanta GA: Centers for Disease Control and Prevention; 2016. URL: http://www.cdc.gov/chronicdisease/overview/ [accessed 2015-07-10] [WebCite Cache ID 6ZvJZqEyM]
9. Mirolla M. The cost of chronic disease in Canada. Ottawa, Ontario: The Chronic Disease Prevention Alliance of Canada; 2004. URL: http://www.gpiatlantic.org/pdf/health/chroniccanada.pdf [accessed 2015-07-10] [WebCite Cache ID 6ZvK7QtQv]
10. Chronic disease indicator framework, 2013 Edition. Ottawa Ontario: Centre for Chronic Disease Prevention, Public Health Agency of Canada; 2013. URL: http://infobase.phac-aspc.gc.ca/cdif/Publications/Quick-Stats_2013_EN_FINAL.pdf [accessed 2015-07-10] [WebCite Cache ID 6ZvKGLymO]
11. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition-multimorbidity. JAMA 2012 Jun 20;307(23):2493-2494 [FREE Full text] [doi: 10.1001/jama.2012.5265] [Medline: 22797447]
12. Sevick MA, Trauth JM, Ling BS, Anderson RT, Piatt GA, Kilbourne AM, et al. Patients with complex chronic diseases: perspectives on supporting self-management. J Gen Intern Med 2007 Dec;22 Suppl 3:438-444 [FREE Full text] [doi: 10.1007/s11606-007-0316-z] [Medline: 18026814]
13. Crossing the Quality Chasm: a New Health System for the 21st Century. Washington DC: National Academy Press; 2001. URL: http://www.nap.edu/read/10027/chapter/1 [WebCite Cache ID 6ZvKsP3zw]
14. Naismith L, Ballem P, Baxter R, Bergman H, Colin-Thomé D, Herbert C, et al. Transforming Care for Canadians with Chronic Health Conditions: Put People First, Expect the Best, Manage for Results. Ottawa, ON: Canadian Academy of Health Sciences; 2010. URL: http://www.cahs-acss.ca/wp-content/uploads/2011/09/cdm-final-English.pdf [accessed 2015-07-10] [WebCite Cache ID 6ZvOHu9D0]

15. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [Medline: 8941260]
16. Wagner EH, Glasgow RE, Davis C, Bonomi AE, Provost L, McCulloch D, et al. Quality improvement in chronic illness care: a collaborative approach. Jt Comm J Qual Improv 2001 Feb;27(2):63-80. [Medline: 11221012]
17. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4 [FREE Full text] [Medline: 10345255]
18. Preventing and managing chronic disease: Ontario's framework. Ottawa, Ontario: Ministry of Health and Long Term Care; 2007. URL: http://www.health.gov.on.ca/en/pro/programs/cdpm/pdf/framework_full.pdf [accessed 2015-07-10] [WebCite Cache ID 6ZvOXEHs2]
19. Price M, Lau FY. Provider connectedness and communication patterns: extending continuity of care in the context of the circle of care. BMC Health Serv Res 2013;13:309 [FREE Full text] [doi: 10.1186/1472-6963-13-309] [Medline: 23941179]
20. Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R. Continuity of care: a multidisciplinary review. Brit Med J 2003 Nov 22;327(7425):1219-1221 [FREE Full text] [doi: 10.1136/bmj.327.7425.1219] [Medline: 14630762]
21. Forrest CB, Glade GB, Baker AE, Bocian A, von Schrader S, Starfield B. Coordination of specialty referrals and physician satisfaction with referral care. Arch Pediatr Adolesc Med 2000 May;154(5):499-506. [Medline: 10807303]
22. Stille CJ. Communication, comanagement, and collaborative care for children and youth with special healthcare needs. Pediatr Ann 2009 Sep;38(9):498-504. [doi: 10.3928/00904481-20090820-09] [Medline: 19772236]
23. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
24. Katz SJ, Nissan N, Moyer CA. Crossing the digital divide: evaluating online communication between patients and their providers. Am J Manag Care 2004 Sep;10(9):593-598 [FREE Full text] [Medline: 15515991]
25. Liederman EM, Lee JC, Baquero VH, Seites PG. Patient-physician web messaging. The impact on message volume and satisfaction. J Gen Intern Med 2005 Jan;20(1):52-57 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40009.x] [Medline: 15693928]
26. Locke KA, Duffey-Rosenstein B, De Lio G, Morra D, Hariton N. Beyond paging: building a web-based communication tool for nurses and physicians. J Gen Intern Med 2009 Jan;24(1):105-110 [FREE Full text] [doi: 10.1007/s11606-008-0827-2] [Medline: 18958533]
27. eHealth Ontario. 2010. ONE Mail Direct: privacy impact assessment summary URL: http://www.ehealthontario.on.ca/images/uploads/pages/documents/ONEMailDirect_PIA_EN.pdf [accessed 2015-12-11] [WebCite Cache ID 6dhepDFj7]
28. Hello Health. 2015. Patient portal: patient engagement URL: https://hellohealth.com/solutions/patient-portal/ [accessed 2015-12-11] [WebCite Cache ID 6gBFdzkf1]
29. Beth Israel Deaconess Medical Center. 2015. PatientSite URL: https://www.patientsite.org/login.aspx [accessed 2015-12-11] [WebCite Cache ID 6dhfnUvJD]
30. Knowledge4You Corp. 2015. Linking Health Professionals URL: http://www.linkinghealthprofessionals.com/ [accessed 2015-12-11] [WebCite Cache ID 6dhfrziCp]
31. Popovici I, Morita PP, Doran D, Lapinsky S, Morra D, Shier A, et al. Technological aspects of hospital communication challenges: an observational study. Int J Qual Health Care 2015 Jun;27(3):183-188. [doi: 10.1093/intqhc/mzv016] [Medline: 25855753]
32. Sunnybrook Health Sciences Centre. 2015. MyChart URL: http://sunnybrook.ca/content/?page=mychartlogin-learnmore [accessed 2015-12-11] [WebCite Cache ID 6dheYW83x]
33. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
34. Tyze Personal Networks. 2015. Tyze connecting care URL: http://tyze.com/ [accessed 2015-12-11] [WebCite Cache ID 6dhQe6GgO]
35. Koronis Inc. 2014. Medlinx: The clinical social network URL: http://medlinx.org/ [accessed 2015-12-11] [WebCite Cache ID 6dhREfS3i]
36. Dalal AK, Dykes PC, Schnipper JL, Bates DW. Transforming the acute care environment: A web-based patient-centered toolkit (abstract). Journal of Hospital Medicine 2014;9 Suppl 2 [FREE Full text]
37. Accountable Care Transactions Inc. 2015. URL: http://act.md/ [accessed 2015-12-11] [WebCite Cache ID 6dhRMcOMY]
38. Bates DW. Health information technology and care coordination: the next big opportunity for informatics? Yearb Med Inform 2015 Jun 30;10(1). [doi: 10.15265/IY-2015-0020] [Medline: 26123913]
39. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Brit Med J 2008;337:a1655 [FREE Full text] [Medline: 18824488]
40. Beyer H, Holtzblatt K. Contextual Design: Defining Customer-Centered Systems. San Francisco, CA: Morgan Kaufmann Publishers; 1998.
41. 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;46 Suppl2:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]

42. Asaro PM. Transforming society by transforming technology: the science and politics of participatory design. Accounting, Management and Information Technologies 2000 Oct;10(4):257-290. [doi: 10.1016/S0959-8022(00)00004-7]
43. Grzyb MJ, Coo H, Rühland L, Dow K. Views of parents and health-care providers regarding parental presence at bedside rounds in a neonatal intensive care unit. J Perinatol 2014 Feb;34(2):143-148. [doi: 10.1038/jp.2013.144] [Medline: 24202296]
44. Stickney CA, Ziniel SI, Brett MS, Truog RD. Family participation during intensive care unit rounds: goals and expectations of parents and health care providers in a tertiary pediatric intensive care unit. J Pediatr 2014 Dec;165(6):1245-1251. [doi: 10.1016/j.jpeds.2014.08.001] [Medline: 25241179]
45. Fetterman DM. Ethnography: Step-by-Step, 3rd Edition. Los Angeles, CA: Sage Publications, Inc; 2010.
46. Holtzblatt K, Wendell J, Wood S. Rapid Contextual Design: a How-To Guide to Key Techniques for User-Centered Design. San Francisco, CA: Elsevier/Morgan Kaufmann; 2005.
47. Muller MJ. Participatory design: The third space in human-computer interaction. In: Jacko JA, Sears A, editors. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Mahwah, N.J: Lawrence Erlbaum Associates; 2003:1051.
48. Boren T, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Profess Commun 2000;43(3):261-278. [doi: 10.1109/47.867942]
49. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [Medline: 10940958]
50. Altheide DL. Reflections: ethnographic content analysis. Qual Sociol 1987;10(1):65-77. [doi: 10.1007/BF00988269]
51. Morgan DL. Qualitative content analysis: a guide to paths not taken. Qual Health Res 1993 Feb;3(1):112-121. [Medline: 8457790]
52. Katz J. The Silent World of Doctor and Patient. Baltimore, MD: Johns Hopkins University Press; 2002.
53. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999 Sep;49(5):651-661. [Medline: 10452420]
54. Wildevuur SE, Simonse LWL. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
55. Gentles SJ, Lokker C, McKibbon KA. Health information technology to facilitate communication involving health care providers, caregivers, and pediatric patients: a scoping review. J Med Internet Res 2010;12(2):e22 [FREE Full text] [doi: 10.2196/jmir.1390] [Medline: 20562092]
56. Kujala S. User involvement: A review of the benefits and challenges. Behav Inform Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
57. Niès J, Pelayo S. From users involvement to users' needs understanding: a case study. Int J Med Inform 2010 Apr;79(4):e76-e82. [doi: 10.1016/j.ijmedinf.2009.06.007] [Medline: 19660983]
58. Davidson JE. Family presence on rounds in neonatal, pediatric, and adult intensive care units. Ann Am Thorac Soc 2013 Apr;10(2):152-156. [doi: 10.1513/AnnalsATS.201301-006PS] [Medline: 23607848]

Abbreviations

AYAC: adolescents and young adults with cancer
CMC: children with medical complexity
EHR: electronic health records
HCP: health care provider
ICT: information and communication technology
PHR: patient-held record
UCD: user-centered design

Edited by G Eysenbach; submitted 31.07.15; peer-reviewed by S Wildevuur, R Valdez; comments to author 05.10.15; revised version received 14.12.15; accepted 19.01.16; published 24.03.16.

Please cite as:
Kurahashi AM, Weinstein PB, Jamieson T, Stinson JN, Cafazzo JA, Lokuge B, Morita PP, Cohen E, Rapoport A, Bezjak A, Husain A
In the Loop: The Organization of Team-Based Communication in a Patient-Centered Clinical Collaboration System
JMIR Human Factors 2016;3(1):e12
URL: http://humanfactors.jmir.org/2016/1/e12/
doi:10.2196/humanfactors.4996
PMID:27025912

Original Paper

Mental Health Technologies: Designing With Consumers

Simone Orlowski[1,2], BSc, BEd, GDipPsych(Hons); Ben Matthews[3], B Eng (Hons), PhD; Niranjan Bidargaddi[1,2], B Eng (Hons), PhD; Gabrielle Jones[4], BHSc; Sharon Lawn[1], BA, Dip Ed, MSW, PhD; Anthony Venning[1], B Soc Sci, B Health Sc Hons, MPsych(Clinical), PhD; Philippa Collin[2,5], BA, PhD

[1]Flinders Human Behaviour & Health Research Unit, Department of Psychiatry, Flinders University, Bedford Park, Australia

[2]Young and Well Cooperative Research Centre, Abbotsford, Victoria, Australia

[3]School of Information Technology & Electrical Engineering, University of Queensland, Brisbane, Australia

[4]School of Medicine, University of Adelaide, Adelaide, Australia

[5]Institute for Culture and Society, University of Western Sydney, Penrith, Australia

Corresponding Author:
Simone Orlowski, BSc, BEd, GDipPsych(Hons)
Flinders Human Behaviour & Health Research Unit
Department of Psychiatry
Flinders University
Margaret Tobin Centre, FMC
Sturt Road
Bedford Park, 5042
Australia
Phone: 61 8 8404 2615
Fax: 61 8 8404 2101
Email: simone.orlowski@flinders.edu.au

Abstract

Despite growing interest in the promise of e-mental and well-being interventions, little supporting literature exists to guide their design and the evaluation of their effectiveness. Both participatory design (PD) and design thinking (DT) have emerged as approaches that hold significant potential for supporting design in this space. Each approach is difficult to definitively circumscribe, and as such has been enacted as a process, a mind-set, specific practices/techniques, or a combination thereof. At its core, however, PD is a design research tradition that emphasizes egalitarian partnerships with end users. In contrast, DT is in the process of becoming a management concept tied to innovation with strong roots in business and education. From a health researcher viewpoint, while PD can be reduced to a number of replicable stages that involve particular methods, techniques, and outputs, projects often take vastly different forms and effective PD projects and practice have traditionally required technology-specific (eg, computer science) and domain-specific (eg, an application domain, such as patient support services) knowledge. In contrast, DT offers a practical off-the-shelf toolkit of approaches that at face value have more potential to have a quick impact and be successfully applied by novice practitioners (and those looking to include a more human-centered focus in their work). Via 2 case studies we explore the continuum of similarities and differences between PD and DT in order to provide an initial recommendation for what health researchers might reasonably expect from each in terms of process and outcome in the design of e-mental health interventions. We suggest that the sensibilities that DT shares with PD (ie, deep engagement and collaboration with end users and an inclusive and multidisciplinary practice) are precisely the aspects of DT that must be emphasized in any application to mental health provision and that any technology development process must prioritize empathy and understanding over innovation for the successful uptake of technology in this space.

(JMIR Human Factors 2016;3(1):e4) doi:10.2196/humanfactors.4336

KEYWORDS

design thinking; participatory design; mental health; technology

Introduction

In light of recent reports that there are almost as many mobile phone subscriptions (6.8 billion) as there are people on Earth (7 billion) [1], more humans are connected and have access to a wide range of information and services than ever before. In the context of this "increased access to information" the promise of the Internet and digital technologies is especially powerful in the prevention and treatment of mental health conditions, an area that has been historically impeded by issues of stigma and misinformation as well as disease-specific, geographical, and financial barriers to help-seeking and service engagement [2-5]. Despite growing interest in the promise of e-mental health preventive/treatment interventions, little supporting literature exists to guide their design and the evaluation of their effectiveness [6-8].

In line with an extensive literature on consumer participation in health care and mental health care more broadly [9-16], human-centered design processes have been identified as a method or set of techniques that assist with good design [17-22]. Both participatory design (PD) and design thinking (DT) have emerged as approaches that hold significant potential for supporting the design of technology-based youth e-mental health and well-being interventions [8,20,23-26]. For example, large-scale PD is embedded within Young and Well Cooperative Research Centre (CRC) [20,27] practice. The CRC combines end-user engagement and youth participation to "explore and understand the role of new and emerging technologies in the lives of young people" [28]. This paper provides a brief background of the evolution DT and PD, where differences in politics and agenda are explored. We then discuss the applicability of PD and DT to design of e-mental health interventions, particularly in the context of application by novice researcher/practitioners. Finally, we present 2 case studies and highlight similarities and differences in process and outcome, mind-set, and emphasis and draw learnings from each to inform design of e-mental health interventions.

Participatory Design in Brief

PD practice has its earliest roots in Scandinavia where it was employed by computer scientists and systems designers initially in industrial workplaces to preserve the autonomy of employees facing significant changes to the organization of their work due to the introduction of new technologies. In this instance, improved outcomes were achieved due to the context-sensitive and future-oriented approach to the design of technological solutions developed by PD practitioners and the methods they used to involve workers in design [29-31]. A fundamental underpinning of Scandinavian PD was democratic participation in proposed changes to work and skill enhancement for workers [31]. One of the reasons PD gained international recognition was that a number of the early and archetypal examples of PD generated far-sighted and innovative solutions. (For example, the graphical user interface that was generated through the UTOPIA project in the early 1980s was clearly ahead of its time.) The methods of end user participation that were developed and shared out of these projects became adopted elsewhere as pathways to innovation—new means of designing successful and user-friendly systems. This gave rise to other more commercial (and less political) forms of PD, particularly in North America, where usability of software and products replaced the focus on workplace democracy [32].

In this Scandinavian context, the practice of PD was characterized by a 3-stage iterative design process aimed at unlocking a users' tacit knowledge: (1) exploration of work; (2) discovery processes; and (3) prototyping. Each of these stages was organized and enacted with users [29]. More recently, variations of PD have been used in a range of contexts for a variety of purposes, with each implementation variously drawing on aspects of its practice (eg, applying PD as a general mind-set for design, or as a method, or adopting individual PD activities as design techniques [33]). PD, or "co-design" as it is called in its broad application, is now practiced within local communities, in companies and organizations, and between companies/organizations and their business partners and/or customers to tackle complex problems and promote innovation and user-centered design [33]. Increasingly PD has been employed in non-workplace contexts [34] by researchers without specific technical or design training as a means of improving the consumer experience in the design of new health interventions [20]. However, there is as yet little evidence as to whether these kinds of consumer participation in the design of new services succeed in improving the efficacy, implementation, and uptake of technology-based interventions [8].

Design Thinking in Brief

Broadly speaking, DT is a term that refers to what designers and design researchers know about successful design processes (the first Design Thinking Research Symposium was held in 1991) [35-37]. In the past decade, however, it has become a term of reference for the mind-set, practices, and methods for generating innovative solutions, taking its starting point from ordinary people's needs. Popularized by prominent design companies such as IDEO, DT has emerged as an articulation of a commercially successful human-centered design process. DT has been defined as "user-centered innovation with a focus on desirability" [38]. And, like PD, it emphasizes participation with and empathy toward users. Increasingly DT has influenced health care design, as well as delivery and training of the workforce [39-44].

DT reinforces the importance of multidisciplinary teams and their ability to generate a diversity of ideas. To harness the best ideas and output, team members are guided by an empathetic mind-set and methods, along with domain-specific knowledge. Naturally, this requires high levels of interpersonal communication. DT's collaborative mind-set is underpinned by a bias toward action, which reinforces quick-and-dirty prototyping and a fail-early-and-often mentality [39,45]. DT is marketed for its ability to be successfully applied by novice practitioners using practical off-the-shelf toolkit [46,47]. DT is often associated with innovation as it attempts to uncover unidentified or unknown needs and offers a specific (and more prescriptive) way forward for the development of interventions that move beyond basic translation of paper-based processes and interventions onto a technology-based platform [33,45,48]. The Stanford d.School Bootcamp Bootleg is one of many

available toolkits and is characterized by 5 design modes: empathize, define, ideate, prototype, and test [47]. Unsurprisingly, these modes neatly overlay the stages, or frameworks, proposed in traditional PD research [20,29]. The design-focused methods and mind-set, detailed in a resource such as the d.School Bootcamp Bootleg, provide an explicit and accessible method for health researchers to become exposed to a design mind-set and the possibility to innovate in circumstances that may be characterized as including incomplete or confusing information, which is often the starting point for intervention researchers.

PD and DT in Health Care

If consumer involvement and/or a human-centered process is rightfully considered to be a part of good intervention design, then it is imperative to develop standards for and document cases of best practice. Hagen et al suggest a framework and techniques/methods for application of PD in a youth mental health intervention design context [20]. The guide articulates possible ways of integrating PD with more traditional evidence-based health research. The same adaptation work has not yet been done with respect to DT. Currently, the notion of applying a set of management processes developed in a commercial business and consulting context to sensitive fields such as youth mental health remains insufficiently interrogated with respect to benefit, risk, and applicability. For example, DT privileges in situ observation of end users to gain knowledge of subjective experience and insights for design. Privacy, confidentiality, and risk concerns make this type of brief observational engagement (by nonmental-health professionals) difficult to achieve in practice.

While the Hagen et al [20] PD framework is practical and accessible, it is unlikely that lay (nontechnical or nondesign) or inexperienced PD researchers would have the specific skill sets necessary to proficiently drive an iterative design process. This skill set in this area of research is particularly important when considering the predominantly consumerist rationale (ie, creating usable, effective, and efficient interventions) cited for employing participatory processes [8]. Sanders' research has argued that the application of PD as a mind-set to guide predesign, discovery, and design initiatives "is best executed by very experienced research practitioners or by young, intuitive practitioners" [33]. This suggests that in the hands of lay and/or inexperienced researchers, PD may risk losing some of its power to create innovative solutions to future problems. This argument suggests a set of learnings and experiences that are tacit in the PD designer-researcher. It is worth emphasizing that while many of the staple PD methods (such as future workshops) appear easy enough to grasp, organize, and conduct, there is a great deal of skill that is required to successfully facilitate them. There is an important distinction between (1) the kinds of tools, processes, and methods used and (2) the mind-set underlying the approach taken. This raises questions around who is best placed to conduct the research and the kinds of interdisciplinary collaborations necessary for successful application of PD in health research contexts.

In contrast, the DT toolkits actively promote, and are arguably intended for, use by novice practitioners. For example, the method cards of a DT resource such as the Stanford d.School Bootcamp Bootleg [47] are deliberately specific in nature and are promoted in such way as to encourage wide dissemination and use. While this may be appealing for inexperienced researchers wishing to adapt design and innovation methods to e-mental health intervention design, it remains unknown just how effective they are in delivering on their promise of scaffolding novice practitioners through a successful design project. The lure of greater innovation in health care, as promised by the DT toolkits, is strong; the requisite skill and practice, however, involved in leading a DT project should not be underestimated, a point clearly highlighted in the following case study.

Case Studies

Beyond the obvious differences in their respective agendas and politics, articulating universal or consistent distinctions between PD and DT practice can be difficult because their similarities are numerous. Both can be categorized under the umbrella term "human-centered design" and are linked to social innovation; collaborative, inclusive, and multidisciplinary practice; and iterative prototyping [31,45,49]. Moreover, DT and PD employ many of the same methods/techniques; for example, they both draw heavily from ethnographic fieldwork methods in their use of interviewing and observation and from design disciplines such as interaction design with techniques such as personas and scenarios [47,50]. Despite these macro similarities, subtle distinctions between the 2 do exist. These distinctions are best made obvious in their practical application; therefore, we present a case study of each to draw these out with the aim of better understanding their applicability to e-mental health and well-being intervention design.

The first case study describes a service design project carried out by an in-house design team at Kaiser Permanente, an American health care provider [51]. Kaiser Permanente is well known for its commitment to innovation and large-scale organizational application of DT [52]. The current case study describes use of DT in redesign of an initial DT service innovation—the Nurse Knowledge Exchange (NKE). This strategy aimed to improve nursing communication and handover (between shifts) in the organization's hospitals. It did this by moving handovers at shift change from the employee breakroom to the patient bedside—a specific example of the type of innovation possible in application of DT. Five years later, the design team was tasked with the redesign of the NKE strategy due to incomplete and inconsistent uptake throughout the organization's hospitals.

In their revision of NKE, Lin et al [51] describe a typical DT cycle—observing and interviewing followed by idea generation/design sessions, prototyping, and field testing. The process, as in most applications of DT, was rapid and expert-led (ie, controlled from start to finish by the design team), and it called on end users, which included staff from all organizational levels but no patients, for contributions at various stages—particularly during interviewing/observing and field testing. The end result was NKEplus.

The authors described heavy resistance to implementation of the NKEplus strategy outside the pilot site, which they attributed

to skepticism in understanding exactly where the solutions that underpinned NKEplus originated. Lin and colleagues believed nurses throughout Kaiser Permanente's hospitals did not see the need for change to their current handover practice and therefore had not bought into the NKEplus strategy. Lin et al [51] highlight that, in their organization, DT-based innovations and change are normally coupled with training support and formal changes to work roles and position descriptions. The rest of the case study details re-implementation of NKEplus, a process that resulted in higher uptake and buy-in for NKEplus organizationwide. This (ultimately more successful) re-implementation process shares a number of similarities with the PD case study, thus 2 case studies are described in parallel in the following section.

The second case study investigates adaptation of PD to a health context. Specifically, it concerns design of an eHealth portal to assist patients undergoing treatment for weight loss [53]. In contrast to the designer-led NKE redesign described above, the authors characterize the process as a design partnership with end users (which in this case were health care professionals and their patients). Moreover, as compared to the DT example, the PD design process took place in a research, not service, context that is typical of their respective applications.

As far as can be determined from the article, Das and Svanaes [53] began the project with a preconceived idea that an eHealth solution could assist patients undergoing weight loss treatment (similar to the DT example in which the overall aim was to improve nursing communication and handovers). Where the process differs from the DT example is that, as per the authors' description, the actual design ideas came from the end users in future workshops that are typical of traditional PD practice. The health care professionals and patients who attended the future design workshops acknowledged the need for support in their treatment via self-help (eg, educational materials, reminders, asynchronous communication between provider and patient, etc) and suggested the possibility of an eHealth portal, which informed the prototypes that were presented to end users in subsequent workshops. The authors also investigated the differing priorities for various end users in the eventual design solution. Moreover, when an existing platform was presented to end users as a possible design solution, it was deemed insufficient and the researchers commissioned the custom build of a product that would meet end users' requirements. This process took a year to complete, which amounts to a much longer timeframe compared to the rapid DT process described above.

In their second attempt to implement the NKEplus strategy, Lin et al [51] employed a more participatory version of DT via a "soft start" implementation process that made space for end user customization of the strategy. In contrast to initial implementation, the soft-start implementation was characterized by participation with "everyone on the same level conversing as peers" in the process. It also highlighted the fail-early-and-often mentality of DT, observable in the quick-and-dirty approach to trialing end-user-generated new ideas. Importantly, the authors ceded control over the solutions developed to the participants; for example, when participants raised concerns or criticisms with the proposed changes (or addressed them to the facilitators), the authors responded by asking the other participants to present how they would recommend that the issue be handled. In this respect, there is a clear priority of the process and quality of participation over specific details of the design outcome. The end result, however, was greater buy-in, more compliance, and improved outcomes for their hospitals. Like the PD case study, this process took significantly longer and, arguably, represented a more realistic process for changing long-standing ways of working (see also Carlgren [52]). The authors note that other teams using DT in their work at Kaiser Permanente had experienced similar disengagement, where the innovations lacked sustainability in sites outside the origin of development. Lin et al [51] note the need for the design to arise out of end users' own concerns, which arguably is the central tenant of DT.

While the Das and Svanaes PD project [53] involved a limited number of end users, there was transparency in the origin of design ideas. The DT and PD teams began with similar processes (eg, interviews, observations) but then diverged, with the PD researchers working with end users in idea generation whereas the DT team did this internally. We are unable, however, to determine whether the more participatory process employed by Das and Svanaes resulted in greater uptake and buy-in by end users with the final implementation; as with much research in PD, the focus of the paper is on how the methods of participation they used elicited valuable insights for design rather than the success of the resulting system in use.

Discussion

The Lin et al [51] case study highlights that DT approaches can be employed in ways that limit the participation of non-designers to expert informants of the contexts of use, or evaluators of ideas, that have been generated through the process. This traditional, less participatory application of DT appears more likely to encounter difficulties and/or resistance in a health care context. The case study contains clear lessons for design of e-mental health and well-being interventions, many of which will be implemented in organizational contexts. Design solutions not generated with end users themselves are more likely to fail, a notion that receives support elsewhere in the literature [38,54]. The manner and method in which design ideas are introduced, discussed, and progressed requires careful consideration for technology design in mental health, a context that is principally composed of highly educated and experienced health professionals who are afforded considerable autonomy in their daily work. Modern application of PD in health intervention research leverages professional and consumer expertise to collaboratively achieve good design outcomes. Its egalitarian mind-set and process may be better suited to mental health professionals who regularly rely on their clinical judgement and expertise in high risk, complex situations. Drawing from and appreciating this experience through meaningful collaboration, as demonstrated in the Das and Svanaes [53] PD project and the more inclusive process of the NKEplus redesign, is likely to yield greater uptake and longevity of research outputs in context. This claim is supported by Lin et al who, along with other DT experts in their organization, report experiencing

ongoing difficulties with bedding down change initiatives that result from traditional expert-led application of DT methods.

One may ask, in promotion and practice of traditional DT methods, are we unhelpfully replacing one expert-led model in health research with another? The difficulty experienced by the DT teams throughout Kaiser Permanente highlight potential inherent limitations in the DT methods for a health care context and the level of experience required for effective practice (or adaptation) of them. The highly experienced team that led this project reported many problems with generating long-term change as a result of the innovation that came out of their DT cycle(s). Furthermore, in selecting the case study for this paper, DT projects in a health care context were scarce and novice-led DT projects were nonexistent. In light of these findings, the claims of novice user uptake of DT seem optimistic at best.

The Das and Svanaes [53] project demonstrates the value of PD for buy-in and uptake of interventions; however, the traditional focus on process over outcome in PD research leaves unresolved questions around its utility as a methodology for intervention design, development, and implementation. From a non-design specialist perspective, the Das and Svanaes paper [51] clearly articulated their methods and techniques, however, the method cards in the DT toolkits more clearly articulate the designer skill set (ie, the tacit mind-sets and capabilities or what to look for and why). For example, the d.School Bootcamp Bootleg [47] articulates mind-sets and behaviors, particularly around empathy and quick-and-dirty prototyping (and show don't tell), which may combine nicely with the participatory, egalitarian elements of PD. In the absence of these designerly mind-sets, it is likely that the early interview and observation work could miss the design perspective and end up an ethnographic study. This is problematic as, while this phase of the design cycle possesses an ethnographic-like quality in that it attempts to better understand existing workflows, circumstances, and people's subjective experience, it should also elicit data around tensions, contradictions, and opportunities for design—crucial design elements that may be overlooked with a purely ethnographic mind-set.

Conclusions

The very clear articulation of mind-set (and output expected from a particular method/technique) in the DT toolkits (such as the progression from empathize to point-of-view to ideate in the early stages of a DT project) provide clarity and design direction for the ethnographic and observational components of design projects. Much can be learned from this approach in health intervention design research and the value of ongoing dialogue and collaboration between health and design research disciplines in this space should not be underestimated. As discussed in the introduction, however, access to mental health workplaces for observation is not an easily negotiated proposition. In comparison to DT, the more integrated nature, and egalitarian purpose, of PD projects supports greater opportunities for meaningful collaboration between research and clinical practice. If the mental health workforce can see the value of the project (because they have played key roles in its origin), research projects stand a greater chance of accessing the individuals and environments they require for intervention design.

We might also note in conclusion that there is a sentiment within the design research community that the notion of design thinking is in danger of being superficially reduced to a toolbox of easy-to-apply methods that appear to offer recipe-like solutions to a vast range of complex problems. This is a serious concern, and it is worth pointing out that the curricula of most studio-based design programs in higher education neither contain nor resemble what has become visible as design thinking. The existence of resources like Stanford's d.School Bootcamp Bootleg, a suite of methods that are freely distributed and packaged in step-by-step instructions is, we believe, a generous gift to the community at large. But their value in application to new and complex spaces (mental health services being our foremost concern in this paper) must be tied to the mind-set in which they are employed. In this domain, such a mind-set ought to draw from both studio-based design disciplines that have given rise to design thinking and from the social and ethical imperatives of participatory design. From design thinking disciplines, such a mind-set incorporates an appreciation of the nature of design as an exploratory, iterative, uncertain, and social form of inquiry (and synthesis) that is never perfect and never quite finished. This understanding of design practice is articulated well in Schön [55]. From participatory design disciplines, the mind-set involves an appreciation that good design emerges from thoughtful and humble facilitation, that participants need to be given the opportunity to take multiple and active roles in all aspects of design, and that shared ownership over proposals for change can be a more valuable form of innovation than technological novelty and disruption. If the design object and/or outcomes require widespread organizational uptake, handing over control of the design process (as in PD) in appreciation of this context can be just as important as the eventual product in generating (and managing) the change.

We in the e-mental health research community must debate and reflect on exactly what we are trying to achieve through the adoption of DT or PD in our work. Do we seek to incorporate new and potentially disruptive ways of working because they are freely available and promise (narrowly defined ideas of) innovation? Or are we in pursuit of methods and interventions that privilege the needs, voice, and contribution of health consumers and professionals? Moreover, from an ethical and moral perspective, egalitarian ways of working such as those exemplified by PD also represent a promising opportunity to redress the legacy of consumer disempowerment in mental health.

Acknowledgments

This research is part of a collaborative project co-funded by the Young and Well Cooperative Research Centre, Flinders University, and Country Health SA. The Young and Well CRC was established under the Australian Government's Cooperative Research Centres Program.

Conflicts of Interest

None declared.

References

1. Wang Y. Time Inc. 2013. More People Have Cell Phones Than Toilets, UN Study Shows URL: http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/ [accessed 2015-01-21] [WebCite Cache ID 6VjuCGysM]
2. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
3. Lambert Tim J R, Newcomer JW. Are the cardiometabolic complications of schizophrenia still neglected? Barriers to care. Med J Aust 2009 Feb 16;190(4 Suppl):S39-S42. [Medline: 19220173]
4. Lawn S. In it together: physical health and well-being for people with mental illness. Aust N Z J Psychiatry 2012 Jan;46(1):14-17. [doi: 10.1177/0004867411432079] [Medline: 22247088]
5. Muir-Cochrane E. Medical co-morbidity risk factors and barriers to care for people with schizophrenia. J Psychiatr Ment Health Nurs 2006 Aug;13(4):447-452. [doi: 10.1111/j.1365-2850.2006.01002.x] [Medline: 16867129]
6. Hilgart MM, Ritterband LM, Thorndike FP, Kinzie MB. Using instructional design process to improve design and development of Internet interventions. J Med Internet Res 2012;14(3):e89 [FREE Full text] [doi: 10.2196/jmir.1890] [Medline: 22743534]
7. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
8. Orlowski S, Lawn S, Venning A, Winsall M, Jones G, Wyld K, et al. Participatory research - one piece of the puzzle: A systematic review of consumer involvement in design of technology-based youth mental health and wellbeingwell-being interventions. JMIR Human Factors 2015;2(2).
9. National HealthMedical Research Council (NHMRC). Canberra: Commonwealth of Australia, 2005. 2005. A model framework for consumer community participation in health medical research URL: http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r33.pdf [accessed 2015-01-25] [WebCite Cache ID 6VpwrEvfL]
10. Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, et al. Systematic review of involving patients in the planning and development of health care. BMJ 2002 Nov 30;325(7375):1263 [FREE Full text] [Medline: 12458240]
11. Boote J, Telford R, Cooper C. Consumer involvement in health research: a review and research agenda. Health Policy 2002 Aug;61(2):213-236. [Medline: 12088893]
12. Boeck T, Collin P. Youth Adult Researcher Reflections on Participatory Research in Australia and the United Kingdom. Fleming J. Boeck T editors. Involving Children and Young People in Health and Social Care Research. Oxford, UK: Routledge 2012:197-208.
13. James AM. Principles of youth participation in mental health services. Med J Aust 2007 Oct 1;187(7 Suppl):S57-S60. [Medline: 17908029]
14. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650. [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
15. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. A systematic review of the impact of patient and public involvement on service users, researchers and communities. Patient 2014;7(4):387-395. [doi: 10.1007/s40271-014-0065-0] [Medline: 25034612]
16. Ennis L, Wykes T. Impact of patient involvement in mental health research: longitudinal study. Br J Psychiatry 2013 Nov;203(5):381-386 [FREE Full text] [doi: 10.1192/bjp.bp.112.119818] [Medline: 24029538]
17. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
18. Coyle D, Doherty G, Matthews M, Sharry J. Computers in talk-based mental health interventions. Interact Comput 2007;19(4):545-562.
19. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. Interacting with Computers 2010 Jul;22(4):243-252. [doi: 10.1016/j.intcom.2010.02.006]

20. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Participatory Design of evidence-based online youth mental health promotion, intervention and treatment. In: Melbourne, Australia: Young and Well Cooperative Research Centre. Melbourne: Young and Well Cooperative Research Centre; 2012.
21. Howe D, Batchelor S, Coates D, Cashman E. Nine key principles to guide youth mental health: development of service models in New South Wales. Early Interv Psychiatry 2014 May;8(2):190-197. [doi: 10.1111/eip.12096] [Medline: 24251956]
22. Anderson JE, Lowen CA. Connecting youth with health services: Systematic review. Can Fam Physician 2010 Aug;56(8):778-784 [FREE Full text] [Medline: 20705886]
23. MacFadyen JS. Design thinking. Holist Nurs Pract 2014;28(1):3-5. [doi: 10.1097/HNP.0000000000000008] [Medline: 24304625]
24. Bernstein R. Drop that pipette: science by design. Cell 2011 Oct 28;147(3):496-497 [FREE Full text] [doi: 10.1016/j.cell.2011.10.010] [Medline: 22036559]
25. Patel MI, Moore D, Blayney DW, Milstein A. Transforming cancer care: are transdisciplinary approaches using design-thinking, engineering, and business methodologies needed to improve value in cancer care delivery? J Oncol Pract 2014 Mar;10(2):e51-e54 [FREE Full text] [doi: 10.1200/JOP.2013.000928] [Medline: 24371302]
26. Johnson AE, Winner L, Simmons T, Eid SM, Hody R, Sampedro A, et al. Using Innovative Methodologies From Technology and Manufacturing Companies to Reduce Heart Failure Readmissions. Am J Med Qual 2014 Dec 15 [FREE Full text] [doi: 10.1177/1062860614562627] [Medline: 25512952]
27. Young and Well Cooperative Research Centre. Australian Government. 2015. URL: http://www.youngandwellcrc.org.au/ [accessed 2015-12-23] [WebCite Cache ID 6e0JuOdla]
28. Young and Well Cooperative Research Centre. Young and Well Cooperative Research Centre Annual Highlights 2012-13. URL: http://www.youngandwellcrc.org.au/wp-content/uploads/2014/03/YW-CRC-Highlights-2012-13-FINAL.pdf [accessed 2015-01-25] [WebCite Cache ID 6VpxwxZaA]
29. Spinuzzi C. Tech Commun; 52(2). 2005. The methodology of participatory design URL: https://faculty.washington.edu/socha/css572winter2012/spinuzzi%20methodology%20of%20participatory%20design%20s5.pdf [accessed 2015-01-08] [WebCite Cache ID 6VPoFnDLi]
30. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: Using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
31. Ehn, P. Scandinavian design: On participation and skill. In: Schuler D, editor. Participatory design: Principles and Practices. Hillsdale, New Jersey: Erlbaum Associates; 1993:41-77.
32. Muller M, Wildman D, White E. Taxonomy of participatory design practices: A brief practitioner's guide. Communications of the ACM 1993;36(4):26-28 [FREE Full text]
33. Sanders EBN. In: Held M, Joost G editors. Wer Gestaltet die Gestaltung? Praxis, Theorie und Geschichte des Partizipatorischen Designs. 2013. Perspectives on Design in Participation URL: http://www.maketools.com/articles-papers/Sanders2013Perspectives.pdf [accessed 2015-01-27] [WebCite Cache ID 6Vsz4mgTa]
34. Wadley G, Lederman R, Gleeson J, Alvarez-Jimenez M. Participatory design of an online therapy for youth mental health. 2013 Presented at: roceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration; November 25 - 29, 2013; Adelaide, Australia p. 517-526.
35. Rodgers PA. Articulating design thinking. Design Studies 2013 Jul;34(4):433-437. [doi: 10.1016/j.destud.2013.01.003]
36. Cross N. Designerly Ways of Knowing: Design Discipline Versus Design Science. Design Issues 2001 Jul;17(3):49-55. [doi: 10.1162/074793601750357196]
37. Dorst K. The core of 'design thinking' and its application. Design Studies 2011 Nov;32(6):521-532. [doi: 10.1016/j.destud.2011.07.006]
38. Wölbling A, Krämer K, Buss C, Dribbisch K, LoBue P, Taherivand A. Design Thinking: An Innovative Concept for Developing User-Centered Software. In: Maedche A, Botzenhardt A, Neer L, editors. Software for People: Fundamentals, Trends and Best Practices. New York: Springer Berlin Heidelberg; 2012:121-136.
39. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
40. Doss H. Forbes. 2014. Design Thinking In Healthcare: One Step At A Time URL: http://www.forbes.com/sites/henrydoss/2014/05/23/design-thinking-in-healthcare-one-step-at-a-time/ [WebCite Cache ID 6VpzEIPY8]
41. Spurrier B. Mayo Foundation for Medical Education and Research. 2015. Design Thinking URL: http://www.mayo.edu/center-for-innovation/what-we-do/design-thinking [accessed 2015-01-25] [WebCite Cache ID 6VpzLPrqg]
42. Brown T. Ideo. 2013. Design Thinking URL: http://designthinking.ideo.com/?P=1201 [accessed 2015-01-25] [WebCite Cache ID 6VpzXLtlL]
43. Arieff A. The New York Times. 2009. A Breath of Fresh Air for Health Care URL: http://opinionator.blogs.nytimes.com/2009/12/13/a-breath-of-fresh-air-for-health-care/?_r=0 [accessed 2015-01-25] [WebCite Cache ID 6Vq08JikE]
44. Ferguson C. Bridgeable. 2012. Applying Design Thinking to Ontario Healthcare URL: http://bridgeable.com/applying-design-thinking-to-ontario-healthcare/ [accessed 2015-01-25] [WebCite Cache ID 6Vq0gCqna]
45. Björgvinsson E, Ehn P, Hillgren P. Design Things and Design Thinking: Contemporary Participatory Design Challenges. Design Issues 2012 Jul;28(3):101-116 [FREE Full text] [doi: 10.1162/DESI_a_00165]

46. IDEO.org. The Field Guide to Human-Centered Design. Canada: IDEO; 2015.
47. Plattner H. d.school, Stanford University Institute of Design. Stanford, CA: Stanford University bootcamp bootleg URL: http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf [accessed 2015-01-25] [WebCite Cache ID 6Vq16FFQD]
48. Owens C, Farrand P, Darvill R, Emmens T, Hewis E, Aitken P. Involving service users in intervention design: a participatory approach to developing a text-messaging intervention to reduce repetition of self-harm. Health Expect 2011 Sep;14(3):285-295. [doi: 10.1111/j.1369-7625.2010.00623.x] [Medline: 20860777]
49. Brown T. Change by design. New York: HarperBusiness; 2009.
50. Muller M, Druin A. Participatory design: the third space in HCI. Hillsdale, New Jersey: L. Erlbaum Associates Inc; 2012.
51. Lin M, Hughes B, Katica M, Zuber C, Plsek P. Service design and change of systems: human-centered approaches to implementing and spreading service design. International Journal of Design 2011;5(2):73-86 [FREE Full text]
52. Carlgren L. Chalmers University of Technology. Sweden Design thinking as an enabler of innovation: Exploring the concept and its relation to building innovation capabilities URL: http://publications.lib.chalmers.se/records/fulltext/185362/185362.pdf [accessed 2015-12-23] [WebCite Cache ID 6e0LZ0vtB]
53. Das A, Svanæs D. Human-centred methods in the design of an e-health solution for patients undergoing weight loss treatment. Int J Med Inform 2013 Nov;82(11):1075-1091. [doi: 10.1016/j.ijmedinf.2013.06.008] [Medline: 23886482]
54. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
55. Schon DA. The reflective practitioner: How professionals think in action. New York: Basic Books; 2001.

Abbreviations

CRC: Cooperative Research Centre
DT: design thinking
NKE: Nurse Knowledge Exchange
PD: participatory design

Edited by G Eysenbach; submitted 10.02.15; peer-reviewed by P Hagen, O Zelenko; comments to author 16.03.15; revised version received 03.07.15; accepted 09.10.15; published 28.01.16.

Please cite as:
Orlowski S, Matthews B, Bidargaddi N, Jones G, Lawn S, Venning A, Collin P
Mental Health Technologies: Designing With Consumers
JMIR Human Factors 2016;3(1):e4
URL: http://humanfactors.jmir.org/2016/1/e4/
doi:10.2196/humanfactors.4336
PMID:27026210

Editorial

Challenges and Paradoxes of Human Factors in Health Technology Design

Plinio P Morita[1*], BEng, MSc, PEng, PhD; Joseph A Cafazzo[1,2,3*], BASc, MHSc, PEng, PhD

[1]Healthcare Human Factors, Techna Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Institute of Biomaterials and Biomedical Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Joseph A Cafazzo, BASc, MHSc, PEng, PhD
Healthcare Human Factors
Techna Institute
University Health Network
TGH/RFE Bldg, 4th Fl.
190 Elizabeth St
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 5514
Fax: 1 416 340 3595
Email: Joe.Cafazzo@uhn.ca

Abstract

Usability testing allows human factors professionals to identify and mitigate issues with the design and use of medical technology. The test results, however, can be paradoxical and therefore be misinterpreted, limiting their usefulness. The paradoxical findings can lead to products that are not aligned with the needs and constraints of their users. We herein report on our observations of the paradox of expertise, the paradox of preference versus performance, and the paradox of choice. Each paradox explored is in the perspective of the design of medical technology, the issues that need to be considered in the interpretation of the test results, as well as suggestions on how to avoid the pitfalls in the design of medical technology. Because these paradoxes can influence product design at various stages of product development, it is important to be aware of the effects to interpret the findings properly.

(JMIR Human Factors 2016;3(1):e11) doi:10.2196/humanfactors.4653

KEYWORDS

equipment design; task performance and analysis; workflow; workload

Introduction

Usability testing is of prime importance in evaluating technology designs. Usability testing can be a powerful tool to validate a design, while equally being useful at identifying flaws. However, when confronted with paradoxical findings, designers and engineers are often left in turmoil over the challenges of interpreting usability test results.

Usability testing as a scientific method is still subject to the limitations of being mostly based on subjective, qualitative evaluations [1]. The data collected during the testing process highly rely on the experience of the experimenter [2], how the experimenter interacts with the participants [3], and the individuals participating in the study [4]. If testing is not done according to established norms of qualitative research, the process ultimately has the potential to result in erroneous findings.

At times, the process of usability testing can also result in some surprising, contradictory, and often-paradoxical findings that may leave human factors professionals (HFPs) perplexed. Only when these paradoxical findings are explained and properly understood by the HFPs can the findings be properly interpreted and the value of the testing be derived in the iterative development process.

In the past decade, hundreds of products have been tested in the usability labs at Toronto General Hospital, part of the University Health Network. During that time, HFPs have routinely identified paradoxical findings on usability tests, which at times, could have led to misinterpretations and erroneous conclusions that in turn could have negatively affected product design.

In this paper, we will explore three paradoxes of health technology design that can confound and mislead both designers and engineers in developing health technologies. These were the most prominent paradoxes identified over the years and the three that could have the most negative impact on design if not accounted for during the evaluations.

The Paradoxes

The Paradox of Expertise: "Do As I Do, Not As I Say"

An iterative, user-centered design (UCD) process of health technology, analogous to any product development, focuses on the use of expert knowledge to identify the requirements, constraints, and features to be included in the final product. Subject-matter experts (nurses, physicians, allied health professionals, among other care providers and patients) are involved in the early stages of product development through interviews and focus groups. Their feedback forms the basis of system's specifications [5-7]. These experts are integrated into the design process because they are considered to bring domain knowledge that is otherwise not available to the design team.

Consequently, as experts, their interactions with the medical technology under development are influenced by their extensive knowledge and well-aligned, mature mental models [8,9]. According to Rasmussen's *Skills, Rules, and Knowledge* framework—which describes why operators with varied levels of expertise and training will behave differently and have distinct psychological processes—these interactions are not necessarily shared by all individuals. These interactions present a unique and refined view of how the expert subset of users interprets the work domain and the interaction with the system [10].

The main premise is that experts will offer greater knowledge in defining product requirements, defining workflows, etc. However, the issue with this approach is that the health technology being developed now reflects only the interactions and constraints of a small percentage of the total users (often, only expert users) who will be interacting with it. The finely tuned mental models of expert individuals are not shared by the majority of the less-experienced users, as described by Hmelo-Silver and Pfeffer [11]. Consequently, the final evaluation only provides a partial view of product specifications because these expert users can have significantly different needs from the health technology being developed. In addition, expert users may have mature mental models that can result in users using shortcuts when interacting with the medical technology and consequently missing important issues with the design. As a result, product specifications identified by experts might significantly differ from the needs of the larger majority of users of the technology.

Some aspects of user interaction design of the system might be left out as a result of the inputs from experts, as their cognitive pathways have allowed them to bypass those components of the workflow. As described by Firesmith [12], "subject matter experts who specify requirements often take certain information for granted and omit it, even though it is not obvious to other stakeholders of the requirement" (p 79). This is the essence of the *paradox of expertise*. In the same direction, the use of expert knowledge can also result in lack of innovation, as experts are usually locked in their own ways, and may demonstrate resistance to innovation.

In our own practice, it has also been observed that there is a very sharp inconsistency and incompleteness between the verbal description of the work performed by experts and how they actually perform their work in the field. This is in alignment with what was discussed by Benner [9] in a previous work where she identifies the difficulty of gathering data from experts and how these expert individuals usually use cognitive shortcuts that they are often not aware of [6,10]. Because their actions and decisions are highly rooted on skill-based behaviors [10], expert's descriptions of the work might be simplistic as they do not fully perceive the wide range of constraints that affect their work. Ultimately, this can potentially lead to distorted representations of the work domain.

The gap between description and performance reinforces the importance of using other ethnographic tools such as in situ observations as part of the requirements gathering and design process as shown in Figure 1 [6,7]. These methods would allow designers to analyze the work domain in situ and gather data without the bias of an expert's limitation. In the requirements gathering stage, designers must ensure that they avoid a distorted representation of the workflow, feature set, and other specifications. It would only be through direct observation that designers could fully comprehend the domain and properly incorporate constraints and requirements into the system.

When designing for complex systems, the lack of complete understanding of the domain can result in a flat information architecture design that leads to a crowded, seemingly complex user interface. Because the designer does not have complete insights into what is important to the user, the final design often lacks the necessary hierarchy of information or functionality that maps to the users' mental model.

For example, in the radiation therapy domain [13], only through proper ethnography were the authors able to identify that the checking procedures during radiation therapy were often skipped because the task was too complex, time consuming, or distractions happened [13]. When asked, professionals would normally state that all checks had been performed. As such, it was necessary for the researchers to be present while tasks were being performed to identify that the skipping had actually happened. This demonstrates that it was only through direct observation that researchers were able to understand the real issues and identify ways to address them. This is a good example of the *paradox of expertise*, where it was important to rely on observed data rather than on verbal reports [13].

The original architecture provided users with the necessary information for the checks, but this information was spread across multiple screens without any logical structure. The authors brought forward items that were previously buried in the interface and difficult for users to locate. By reorganizing the information architecture and through forcing functions in the form of simplified automated checklists, the authors were able to significantly improve the checking process and patient safety. The new interface, when compared with the original

one, showed improved error detection rates and high user satisfaction [13].

Design decisions must be made based on a combination of user-reported data and observed data to ensure that the system is designed for how users actually use it, instead of being designed to how they think they would use it. Although design requirements might be gathered through expert interviews and focus groups, only though the use of observational techniques can designers have a rich understanding of the work domain and the system's hierarchy of information.

Figure 1. Human factors expert embedded in an operating room environment at University Health Network, gaining a deeper understanding of how clinicians actually work.

The Paradox of Preference Versus Performance: How Could Someone Like Something They Cannot Use?

One would expect that when evaluating two possible designs, users would prefer the design in which they had greater success during testing. Oddly, that is not always the case, leaving the HFPs to conclude that the testing was somehow flawed, or they just disregard that user's opinion altogether.

How could someone like something they cannot use?

Contrary to these paradoxical findings, Nielsen and Levy [14] described a positive correlation between user preference and user performance showing that, in general, users prefer systems in which they also performed the best. However, the same authors also argue that there are still many cases in which users prefer systems in which they perform worse. Although users are described to prefer situations in which preference and performance align [14,15], we have identified cases over years of product testing to consider these paradoxical findings as a risk.

As design methods have evolved, more approaches have been made available to influence user behavior by making simple changes in the aesthetics of the device or by using a seemingly novel and engaging control interface. New features might drive users to prefer a particular design simply due to increased affinity for that experience.

Powerful persuasive design can be used to guide how users perform certain tasks, influence user interaction, and drive user behavior. Similarly, design techniques can be used to capture users' attention and persuade them to react positively to a design, which could be flawed or create negative outcomes [16]. Seemingly novel features and a more aesthetically pleasing

design of health technology may drive users' preference, but these do not necessarily result in better task performance.

In practice at Toronto General Hospital's usability labs (Figure 2), cases have been observed where the color palette of a device had a greater influence on nursing preference than on its usability. In this case, the observation of the paradox was further reinforced by the novel user interface of a scroll wheel that nurses found interesting and engaging to use, but did not result in successfully completing tasks. User preference, evaluated through questionnaires, demonstrated that nurses preferred the new device design. Observational and performance data, however, showed that their performance was suboptimal. Besides, the new design led to numerous errors, operational difficulties, and failure to complete tasks. The scroll wheel and the color selection corresponded to design features that can become too salient and lead users to preferring a certain device. For that reason, our team ensures that design evaluations not only rely on the self-reported, subjective opinion of the users, but also on the unbiased, direct observation of their performance.

The paradox of preference versus performance described herein demonstrates the potential of design in affecting user preference, sometimes at the expense of the system's usability. While interpreting the results of such testing, one must be cognizant not to bias his/her conclusion in favor of a design that in the end could be compromised.

Figure 2. Usability labs at Toronto General Hospital showing a complete set up of a simulated operating room (including a patient simulator).

The Paradox of Choice: Less Is Often More

A number of studies have demonstrated how choice influences our buying decisions, selection of services, and ultimately how choice impacts our lives [17]. Choice consists of a mental decision-making process in which individuals have to judge merits among a range of options available and select one [18,19]. Although rooted in individual cognitive processes, extensive work over the years has been carried out in understanding how to influence choice by manipulating the access to information and how information is presented to individuals, with regard to marketing, interface design, and product design.

Although choice is often praised as being necessary for proper decision making, extreme situations can result in indecision and discomfort [17]. Schwartz [17] describes how excessive choice has impacted us as individuals and collectively as a society. Especially relevant here is his description of situations in which too much choice for individuals can potentially result in conditions in which a user makes poor choices, or no choice at all.

Within a health care perspective, designers can have the misconception that including more features in a product would be beneficial to caregivers and patients, who would now have a wider range of functionality and operational modes to use and more features to tailor their care. The pitfall is that, by including

those additional features, one can lead caregivers to make poor choices, as described by Schwartz [17].

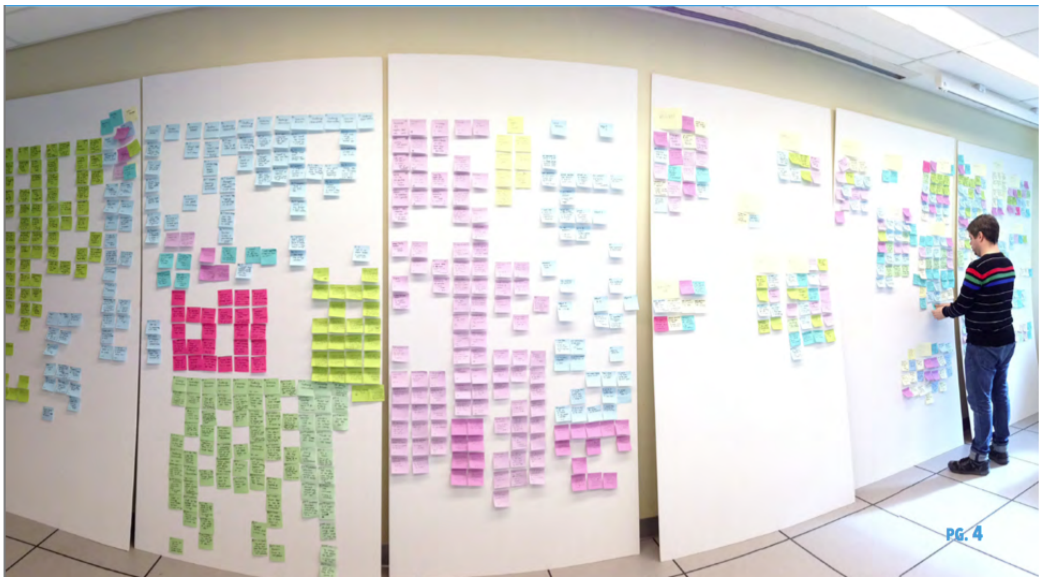
Health care is a highly demanding work environment where caregivers are generally under extreme pressure, which is a perfect situation for excessive choices to become overwhelming and a nuisance at a minimum, and safety hazard at its worst. Adding more choice and options to a single user interface can create uncertainty and distraction to the user. The complexity can create visual noise generated by the new features and cause users to be less efficient, make use errors, and generally provide them with a poorer user experience.

Within a health information technology domain, our teams have observed the effects of the paradox of choice consistently in the design and evaluation of electronic medical record (EMR) systems. To satisfy all possible end users from different specialties and different areas of a health care institution, designers include many features, functions, and information fields on a single user interface. EMR systems are renowned to overload users with choices on a single screen, creating a situation where users struggle to find the necessary information, function, or feature [20,21]. The consequence is that users now have to dig through numerous screens and tabs to find or enter the necessary information, leading to decreased performance, increased frustration, and unnecessary workload. EMR manufacturers have taken a one size-fits all solution that can severely impact the usability of the systems. Hence, EMR manufacturers must be aware of the paradox of choice to design future EMR systems that rely on simplified interfaces that present the user with a limited number of choices, facilitating access to information and reducing load on the user.

We need to be cognizant, however, that health care institutions fail to design their work environment for simplicity of workflow and standardization. Each health care institution prides itself for being unique. Consequently, manufacturers of health care technology have to navigate this complex environment and constantly make critical decisions: design a simple system to the specification of a few organizations or a complex system that fits most organizations? Nonetheless, engineers and designers must be aware of the *paradox of choice*, as during their effort to create a product that satisfies a greater audience, they may end up with an unusable product, which is often the case in EMR systems. The systemic issue with the lack of standardization must be addressed in the long run to ensure that medical technology can be properly designed to maximize benefits and mitigate usability issues. Health care must strive to harmonize their work environment and policies to increase standardization and consequently, facilitate the design of better technology.

This is not to suggest that only extremely simple systems with basic functionality are viable. A delicate balance needs to be cast where designers should aim for an interface in which users are not overloaded with excessive choices, while being inclusive enough to incorporate necessary features for proper operation of the system for the advanced user. Such systems can only be achieved through a detailed and careful design process that incorporates the needs and constraints of the final users (Figure 3).

Figure 3. Usability labs at Toronto General Hospital, where we show the data analysis process through concept mapping and affinity diagrams.



In the End, What Is Necessary for a Successful Design?

Usability testing and other HF methods are undoubtedly useful and powerful tools in the design process. However, one needs to be mindful of the pitfalls discussed here when designing systems and when evaluating the data collected through testing, as they may significantly influence the final design of a health technology. The paradoxes described in this article have the potential of skewing the understanding of the work domain and product requirements by presenting the designers with an incomplete and biased perception of the task. To design a product that is in alignment with the needs of its final users, designers must be aware of the paradoxes of expertise, preference versus performance, and choice, to ensure that their effect on product design is controlled or even mitigated.

The lesson to be learned from the paradoxes described in this paper is that to design health technology aligned with the needs of its final users, engineers and manufacturers must incorporate a gamut of UCD methods (Figure 4) in the design process to gain a comprehensive and realistic understanding of the work domain and user constraints. Observational methods such as cognitive walkthroughs and usability testing provide an opportunity to gather information about how users actually use the technology. The data gathered through these two methods can help minimize the impact of the *paradox of expertise* and the *paradox of preference versus performance*, allowing designers to focus on tailoring the technology based on unbiased usage data. Other methods such as interviews and concept mapping can be used to address the effects of the *paradox of choice*, creating opportunities for designers to identify the needs of each health care professional and organize the requirements into a manageable and tailored version of the technology.

A combination of methods is always necessary to ensure that the system being designed aligns with user needs and works toward bridging some of the gaps identified. Only then it is possible to focus on designing simple and tailored health technology that maximizes benefits to the users without overloading them with choice.

Figure 4. Examples of human factors methods used by human factors professionals at the University Health Network for designing and testing medical technology. Starting from the top left, clockwise, we showcase examples of interviews, cognitive walkthroughs, concept mapping, and usability testing.

Conflicts of Interest

None Declared.

References

1. Dicks RS. Mis-Usability: On the uses and misuses of usability testing. New York: ACM; 2002 Oct Presented at: 20th Annual International Conference on Computer Documentation; Oct 20-23, 2002; Toronto, Canada p. 20-23.
2. Holleran PA. A methodological note on pitfalls in usability testing. Behav Inform Technol 1991 Sep;10(5):345-357. [doi: 10.1080/01449299108924295]
3. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
4. Virzi R. Refining the test phase of usability evaluation: How many subjects is enough? Hum Factors 1992;34(4):457-468. [doi: 10.1177/001872089203400407]
5. Madrigal D, McClain B. Subject-Matter Experts: Putting Users at the Center of the Design Process. 2011. URL: http://www.uxmatters.com/mt/archives/2011/02/subject-matter-experts-putting-users-at-the-center-of-the-design-process.php [accessed 2016-02-09] [WebCite Cache ID 6YS2RS2Az]
6. Vicente KJ. Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Mahwah, NJ: Lawrence Erlbaum Associates; 1999.

7. Beyer H, Holtzblatt K. Contextual Design: Defining Customer-Centered Systems. San Francisco, CA: Morgan Kaufmann; 1998.
8. Gentner D, Stevens AL. Mental Models. New York, NY: Psychology Press; 2014.
9. Benner P. From novice to expert. Am J Nurs 1982 Mar;82(3):402-407. [doi: 10.2307/3462928]
10. Rasmussen J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans Syst Man Cybern 1983 May;SMC-13(3):257-266. [doi: 10.1109/TSMC.1983.6313160]
11. Hmelo-Silver C, Pfeffer NG. Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. Cogn Sci 2004 Feb;28(1):127-138. [doi: 10.1016/S0364-0213(03)00065-X]
12. Firesmith D. Specifying good requirements. J Object Technol 2003;2(4):77-87. [doi: 10.5381/jot.2003.2.4.c7]
13. Chan AJ, Islam MK, Rosewall T, Jaffray DA, Easty AC, Cafazzo JA. The use of human factors methods to identify and mitigate safety issues in radiation therapy. Radiother Oncol 2010 Dec;97(3):596-600. [doi: 10.1016/j.radonc.2010.09.026] [Medline: 21044802]
14. Nielsen J, Levy J. Measuring usability: Preference vs. performance. Commun ACM 1994;37(4):66-75 [FREE Full text] [doi: 10.1145/175276.175282]
15. Moss GA. Psychology of performance and preference: Advantages, disadvantages, drivers and obstacles to the achievement of congruence. J Brand Manag 2007 Apr;14(4):343-358. [doi: 10.1057/palgrave.bm.2550075]
16. Berdichevsky D, Neuenschwander E. Toward an ethics of persuasive technology. Commun ACM 1999;42:51-58. [doi: 10.1145/301353.301410]
17. Schwartz B. The Paradox of Choice: Why More Is Less. New York, NY: Harper Collins; 2009.
18. Shafir E, Simonson I, Tversky A. Reason-based choice. Cognition 1993;49(1-2):11-36. [doi: 10.1016/0010-0277(93)90034-S]
19. Janis IL, Mann L. Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment. New York, NY: Free Press; 1977.
20. Ludwick DA, Doucette J. Primary care physicians' experience with electronic medical records: Barriers to implementation in a fee-for-service environment. Int J Telemed Appl 2009;2009:1-9. [doi: 10.1155/2009/853524]
21. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010;10:231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]

Abbreviations

EMR: electronic medical record
HFP: human factors professional
UCD: user-centered design

Edited by G Eysenbach; submitted 11.05.15; peer-reviewed by S Kitsiou, A Easty, N Wayne, D Pan; comments to author 18.08.15; revised version received 31.10.15; accepted 25.11.15; published 01.03.16.

Please cite as:
Morita PP, Cafazzo JA
Challenges and Paradoxes of Human Factors in Health Technology Design
JMIR Human Factors 2016;3(1):e11
URL: http://humanfactors.jmir.org/2016/1/e11/
doi:10.2196/humanfactors.4653
PMID:27025862

Original Paper

Sociotechnical Human Factors Involved in Remote Online Usability Testing of Two eHealth Interventions

Lori M Wozney[1*], PhD; Pamela Baxter[2*], RN, PhD; Hilary Fast[3*], MA; Laura Cleghorn[2*], MA; Amanda S Hundert[1*]; Amanda S Newton[3*], PhD

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

[2]School of Nursing, Health Sciences Centre, McMaster University, Hamilton, ON, Canada

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

[*]all authors contributed equally

Corresponding Author:
Lori M Wozney, PhD
Centre for Research in Family Health
IWK Health Centre
5850—5980 University Avenue
Halifax, NS, B3K 6R8
Canada
Phone: 1 902 470 7541
Fax: 1 902 470 7542
Email: loriwozney@gmail.com

Abstract

Background: Research in the fields of human performance technology and human computer interaction are challenging the traditional macro focus of usability testing arguing for methods that help test moderators assess “use in context” (ie, cognitive skills, usability understood over time) and in authentic “real world” settings. Human factors in these complex test scenarios may impact on the quality of usability results being derived yet there is a lack of research detailing moderator experiences in these test environments. Most comparative research has focused on the impact of the physical environment on results, and rarely on how the sociotechnical elements of the test environment affect moderator and test user performance. Improving our understanding of moderator roles and experiences with conducting “real world” usability testing can lead to improved techniques and strategies

Objective: To understand moderator experiences of using Web-conferencing software to conduct remote usability testing of 2 eHealth interventions.

Methods: An exploratory case study approach was used to study 4 moderators’ experiences using Blackboard Collaborate for remote testing sessions of 2 different eHealth interventions. Data collection involved audio-recording iterative cycles of test sessions, collecting summary notes taken by moderators, and conducting 2 90-minute focus groups via teleconference. A direct content analysis with an inductive coding approach was used to explore personal accounts, assess the credibility of data interpretation, and generate consensus on the thematic structure of the results.

Results: Following the convergence of data from the various sources, 3 major themes were identified: (1) moderators experienced and adapted to unpredictable changes in cognitive load during testing; (2) moderators experienced challenges in creating and sustaining social presence and untangling dialogue; and (3) moderators experienced diverse technical demands, but were able to collaboratively troubleshoot with test users.

Conclusions: Results highlight important human-computer interactions and human factor qualities that impact usability testing processes. Moderators need an advanced skill and knowledge set to address the social interaction aspects of Web-based usability testing and technical aspects of conferencing software during test sessions. Findings from moderator-focused studies can inform the design of remote testing platforms and real-time usability evaluation processes that place less cognitive burden on moderators and test users.

(JMIR Human Factors 2016;3(1):e6) doi:10.2196/humanfactors.4602

KEYWORDS

usability inspection; Web-conferencing; telemedicine; HCI design and evaluation methods; walk-through evaluation; real-time systems; human centered-computing

Introduction

Traditional usability testing sessions for Internet-based (ie, eHealth) interventions focus on assessing the effectiveness, efficiency, and learnability of and user satisfaction with the intervention. These test situations typically involve in-person, lab-based, or field sessions [1-3]. While a lab setting allows for more experimental control and collection of various types of data during usability testing, it lacks the realism of a field setting. It also precludes deeper engagement from potential end users offered through field-based moderation. Usability moderators leverage many of the skills qualitative researchers already have (eg, building rapport, probing for clarity, getting below top-of-mind responses). To a large extent, successful usability testing depends on the skills of the person moderating the test, which Dumas and Loring suggest is "easy to do, hard to do well" [4].

Fieldwork that includes remote Web-based moderation is a novel approach to usability testing that could potentially mitigate some of the common problems experienced in lab-setting facilitation (eg, cost of maintaining a lab, less authentic or "real world" use contexts) [5]. However, any challenges experienced in computer-mediated communication between test users and moderator has a direct impact on the quality and accuracy of research findings and subsequent decisions about design. For testing of Internet-based interventions aimed at individuals with medically complex situations (eg, comorbidities, chronic illness, and/or frequent relapse cycles), the moderators' ability to confidently use tools that help explore and communicate "use in context" experiences are critical to successful design [6].

Technologies that support real-time, remote collaboration have expanded usability testing possibilities to include geographically remote testing through Web-based moderation (eg, Morae and UserZoom remote usability testing platforms). During Web-based usability testing, the moderator and test user can be geographically separated but can still observe, prompt, and respond to questions in real-time. This approach may help to address the study of more complex eHealth interventions and difficulties that can arise from lab-based and other forms of field-testing when target users are: (1) needed from within a certain clinical population that is geographically dispersed; (2) have limitations in functioning and accessibility due to illness, often the reason for which the intervention was developed; and/or (3) are part of at-risk or age-sensitive groups (eg, minors who would be in school during typical "business hours") that face challenges in travel, time, and cost of attending in-person lab testing [7]. Importantly, studies comparing lab-based testing with remote testing have consistently found no significant difference in usability performance results [8,9].

Web-based usability testing can involve synchronous (ie, moderators and test users are in same place [virtual or physical] at the same time), asynchronous (ie, automated, no real-time interaction), and blended (ie, asynchronous and synchronous) approaches. Remote, synchronous methods are proposed to be useful for usability testing early in the intervention development process. Real-time discussions between the moderator and user can be used to identify usability concerns while prototypes and user interface models are still under development [10] and may potentially save on development costs. While asynchronous, automated methods enable access to large data pools, the reliability of this testing approach has been questioned [11], and it is proposed that this approach may be more time-consuming for the novice tester and result in fewer usability problems being identified [8]. Automated testing methods alone are also not conducive to identifying what Andrezejczak [12] calls the "softer" subjective usability elements (eg, user preferences, misconceptions, underlying values, context variables, motivational attributes, affective attributes) that are better explored through synchronous inquiry methods with a moderator [13].

Web conferencing software packages (eg, GoToMeeting, Cisco WebEx, Microsoft NetMeeting or Lotus Sametime, Blackboard Collaborate, Adobe Connect Pro) are one option for remotely connecting with test users. Although the literature supporting Web-conferencing tools for online collaboration in higher education is extensive [14,15], published research on the use of these tools in moderating usability testing is limited. With the range of functionality provided by these software systems, there is potential to support a diverse range of remote usability session configurations and testing tasks. To date, published research has only begun to explore the role of social environment (ie, individuals present during testing) or the interactions between physical and social environments in usability testing. Study of physical usability test environments suggests that social context plays a substantial role in the quality of usability evaluation results [16,17]. Evaluator effect has been probed by van den Haak and de Jong [18] and interactions between test monitor and test users across multiple in-lab test scenarios were shown to have a significant effect on usability results. The detection of problems and selection of priority usability issues are subject to considerable individual variability [2]. These facets may be equally prominent during Web-based moderation, but there is little research exploring remote usability testing from the moderator's point of view.

The purpose of this study was to: (1) understand moderator experiences using Web-conferencing software in the context of conducting remote usability testing; (2) compare and contrast moderator experiences using the same Web-conferencing software for 2 different Internet-based eHealth interventions; and (3) highlight important practical human-computer interactions qualities that may impact usability testing processes for other researchers.

Methods

Research Design and Usability Testing Context

A single case study approach was used to study the experiences of 4 moderators on 2 projects involving usability testing on eHealth interventions designed and delivered via a "smartsite" software platform called IRIS (intelligent research and intervention software) [19]. This approach allowed for rigorous exploration of the phenomena incorporating multiple perspectives and the dynamism of observations across time and projects [20,21]. Project 1 was an Internet-based anxiety treatment program for adolescents with anxiety disorders. Project 2 was an Internet-based intervention for caregivers of children with fetal alcohol spectrum disorders (FASD). Usability testing was conducted to improve the interventions in terms of the content (ie, therapeutic message, sequence of modules), aesthetics (ie, "look and feel," appropriateness of images), and IRIS platform functionality (eg, customization abilities, site navigation tools, communication features). Usability testing protocols for both projects were approved by institutional research ethics boards and test users provided informed consent. A comparative summary of usability testing set up for the 2 projects is described in Table 1. Figures 1 and 2 illustrate the different usability test scenarios and roles for both projects.

Table 1. Comparison of usability project moderation.

Protocol Feature	Cycle	Project 1 (Anxiety)	Project 2 (FASD)
Number of cycles		2 (same group of test users for both cycles)	2 (new group of test users in each cycle)
Number of test users per cycle, n	Cycle 1	9 (4 youth, 5 clinicians)	10 (4 caregivers, 6 clinicians/health care professionals)
	Cycle 2	8 (4 youth, 4 clinicians)	8 (4 caregivers, 4 clinicians)
Dates of session	Cycle 1	June–July 2013	August–September 2013
	Cycle 2	September 2013	October–November 2013
Number of remote moderators in each session		2	1
Access to intervention prior to remote usability session		No	No
Software version		Blackboard Collaborate 9.7	Blackboard Collaborate 12.5
Average length of usability testing session		133 minutes[a]	62 minutes[b]
Location of moderator(s)		Ontario, Alberta	Nova Scotia
Estimated training time required for moderators to set up usability sessions		40 hours	20 hours
Location of test users		Nova Scotia, Alberta, British Columbia	British Columbia, New Brunswick, Alberta, Saskatchewan, Yukon, Ontario, Manitoba, Northwest Territories
Moderator(s) had prior experience as user in Web conferencing		Yes	Yes
Moderator(s) had prior experience moderating via Web conferencing		No	No
Moderator(s) had prior experience facilitating usability testing		No	No
Moderator(s) had prior experience in facilitating research interviews		Yes	Yes

[a]133 minutes=average time for Cycle 1 and Cycle 2 combined

[b]62 minutes=average time for Cycle 1 and Cycle 2 combined

During the usability testing, moderators used Blackboard Collaborate, a Web-conferencing system and one of the most advanced computer-mediated communication platforms on the market. The system was selected due to its low bandwidth, which accommodates slower user connection speeds making it more widely accessible for test users involved in the 2 projects. Interacting through the Blackboard Collaborate system is designed to mimic face-to-face contexts. Moderators and test users can share screens, indicate a desire to talk by clicking on a "raise hand" button, chat through instant messaging, and "draw" on the virtual whiteboard. The session moderator retains control of the various system tools, but he or she can share that control with others [22]. In addition to the functionality provided, Blackboard Collaborate was selected because it was an institutionally adopted tool at both main research institutions involved in the study, meaning no additional software licensing

fees were required and there was technical support available on-site.

To prepare for using Blackboard Collaborate in the test sessions, moderators viewed demonstration videos, attended online tutorials, participated in mock sessions, and undertook several iterations of trial-and-error. To explore "use in context," both projects configured test sessions to support blended usability testing techniques: "cognitive walk-through" (eg, user is given a task and the evaluator observes user's intentions and the feedback provided by the system's interface); "think-aloud" (ie, "novice" users verbalize their experiences as they work through tasks); and post-hoc interviews and self-report questionnaires.

Figure 1. Web-conferencing test environment setup for ehealth Project 1. Moderator 1 controls recording and access privileges to test environment. Test User and Moderator 2, each in different geographic locations, act as â€œattendeesâ€ with different roles.

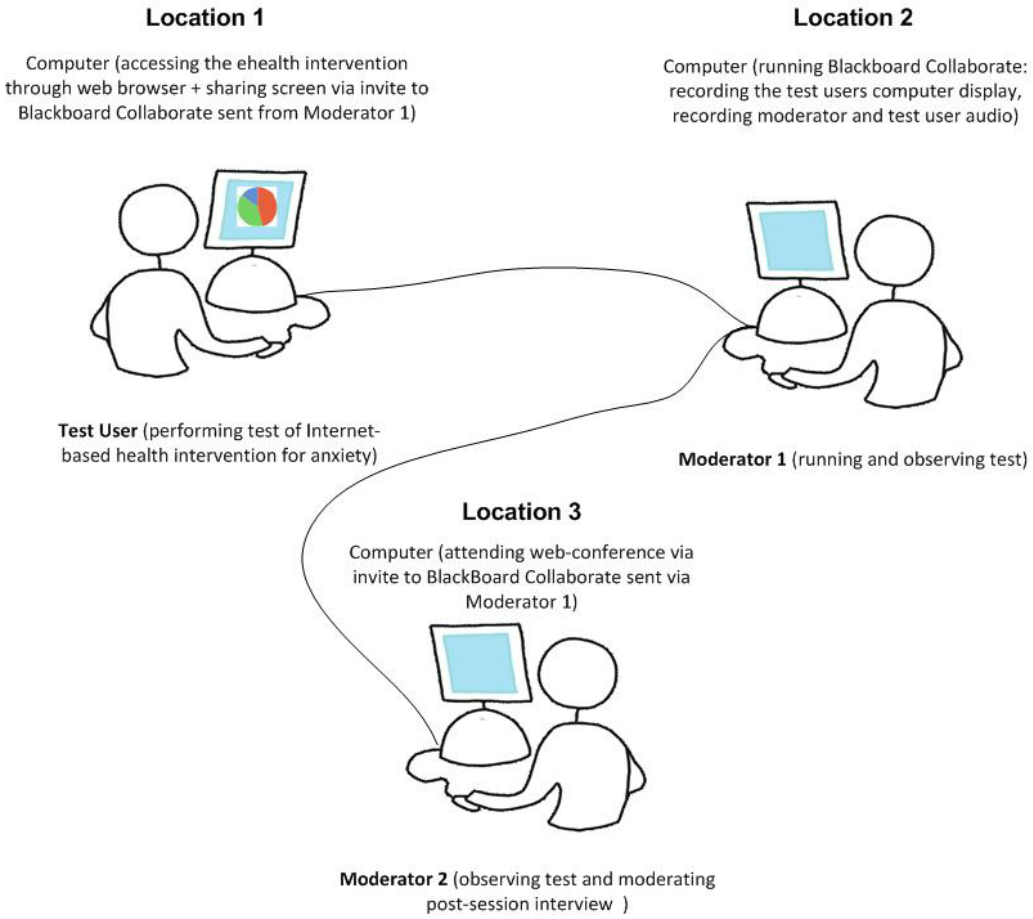
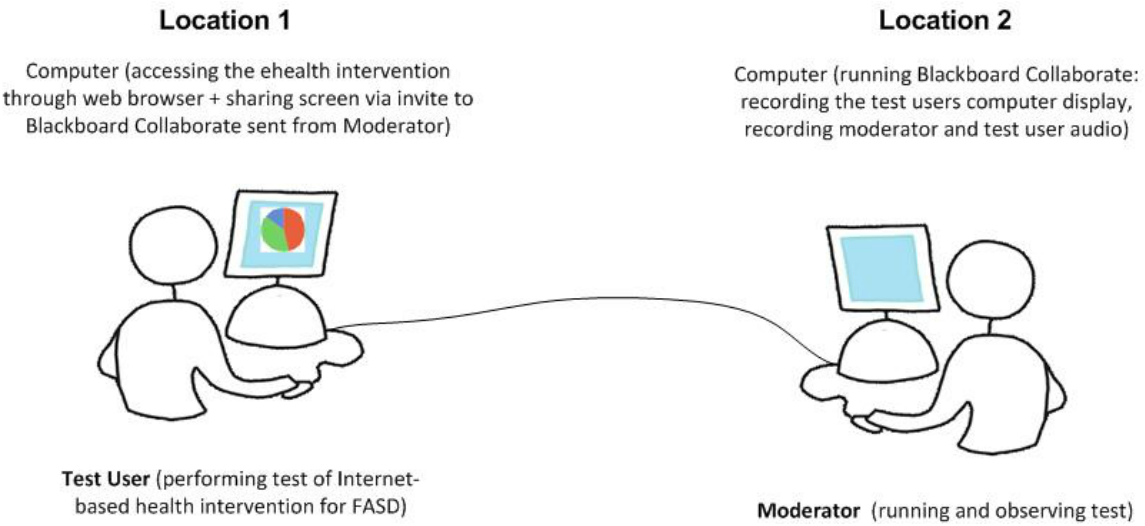


Figure 2. Web-conferencing test environment setup for ehealth Project 2. Moderator controls recording and access privileges. Test User â€œattendsâ€ the web-conference and shares screen so Moderator can observe actions.



Data Collection

Prior to the first test session, session moderators sent an email to test users that provided technical instructions (eg, updating Java, testing audio) and study procedures. Test users in both projects were mailed USB headsets with noise-canceling microphones, if needed. Once logged into the session moderators had to walk users through "auto tuning" tests to ensure they were able to hear and be heard during the session. Test users "shared" their desktop so the moderator could observe them. Although the simultaneous camera feature was available, and offered to test users, none of the test users self-selected to employ this feature. Figure 1 shows an example of the moderator Web-conferencing environment for a Cycle 1 usability test session for Project 1. Project 1 involved 2 moderators in all but 1 test session. The first moderator focused on facilitating the main usability session walk-through tasks. The second moderator facilitated the post-session, open-ended question portion of the test session as more of an interviewer. A third moderator was present for 1 test session only as an observer to be able to provide feedback on usability test processes for the other moderators. In Project 2, only 1 moderator was present during all of the sessions. Moderators in both Projects 1 and 2 made detailed notes during each test session and created summary reports of key observations immediately following each test. Audio and/or video files for all sessions were recorded and saved as JAR files.

Focus groups via teleconference were held with all moderators at 2 points in time. The first was held 2 weeks after Cycle 2 usability test sessions for both projects were complete. Although an immediate debrief would have been ideal, coordinating a multi-site, multi-time-zone research team presented certain scheduling challenges. The second focus group was held 4 weeks later to allow time for incorporating feedback and review. Teleconferences were not digitally recorded, but detailed notes were taken by the first author. Notes included some verbatim statements and paraphrases of verbal statements. The first 90-minute focus group focused on a micro perspective of the data with each moderator describing their personal account and experiences. The over-arching exploratory question being: "What are moderators' experiences using Web-conferencing for conducting remote usability testing?" An additional list of probative questions was circulated to all moderators 1 week prior to the initial focus group. The following list of questions was informally used as the focus group discussion guide:

1. How did Blackboard Collaborate support/hinder you as a researcher? Our team?
2. How do you think the tool supported/hindered our research participants?
3. Was there anything about the tool that surprised you? Really confused you?
4. Which testing activities (cognitive walk-through/think-aloud) was Blackboard Collaborate more useful for? Why?
5. How did it feel for you to be remote from the user/mediated by the computer?
6. Were you concerned about not having any visual cues, such as body language, to guide you? Why?
7. What (if any) ethical issues did you have with using this tool?
8. How do you think using this tool differed from what you would have done face to face?
9. Do you think you captured different types of data using this tool? If yes, in what way?
10. If another researcher was thinking about conducting remote usability sessions would you recommend this tool? Why? Why not?

The second 90-minute focus group occurred following the preliminary data analysis stage as part of planned member check-through debriefing and respondent validation as recommended by Koelsch [23]. During the second focus group, the first author guided the discussion toward theory development from a macro perspective (ie, exploring meaning of collective experiences). The discussion focused on: (1) assessing the credibility of preliminary data interpretation, (2) refining the proposed thematic structure, and (3) evaluating the suitability of examples appearing within the master list of themes [24]. Detailed notes were again taken. Member checks also occurred informally over several weeks during the normal course of observation and conversation with research team members over email, by phone, and in person.

Results

We conducted an iterative thematic analysis whereby data were analyzed from all sources; the analysis was examined and reorganized, the reorganized data was synthesized, and the synthesis was then interpreted [25]. This inductive analytic approach strengthens the reliability of qualitative research [26]. In the first phase, features of each moderator's experience were carefully detailed by a close reading of session transcripts, moderator session notes, and notes from the first focus group. A master list of emergent themes was drafted by the first author after the first focus group and circulated via email to all moderators to promote retrospection and exploration into any issues with the trustworthiness (ie, dependability, confirmability) of the synthesized master list. Written responses to this member check were returned by each moderator via email with suggested changes or clarifications integrated into the draft version. The second phase of interpretation involved exploring convergences and divergences within and between individual accounts. During the second focus group, points that were identified during member check that required clarification were discussed. This second iterative analysis phase allowed us to capture any interesting relationships, patterns, surprises, and inconsistencies among people and within and across sites [27]. Notes from the second focus group were used to further refine master theme examples and descriptions.

Themes

Three major themes emerged from the converged data: (1) experiencing and adapting to unpredictable changes in cognitive load; (2) experiencing challenges in creating and sustaining social presence and untangling dialogue; and (3) collaboratively troubleshooting diverse technical needs and issues with test users. Moderators' experiences were, overall, characterized by generally positive feelings and attitudes toward the experience

of moderating usability testing remotely via Blackboard Collaborate. There was considerable congruence between themes, more so in some cases than others. The interdependence of themes makes it difficult to separate some examples into component parts. Test users shared similar experiences, and although there were idiosyncrasies that marked individual experience, differences between test users were generally characterized by the intensity or depth of these shared experiences.

Theme 1: Moderators Experienced and Adapted to Unpredictable Changes in Cognitive Load

Moderators all agreed that the range of communication features available in Blackboard Collaborate helped support interpretive aspects of usability testing. Test users not only identified problems and errors, but were also able to participate in impromptu interpretation of what problems meant or how they might be solved.

> *I was surprised at how often youth would stop and offer suggestions about how to improve things...they didn't just point out problems...they had a lot of creative ideas about how to improve things...we could brainstorm together.* [MOD 1]

Often these spontaneous interpretive interactions between test users and moderators happened because moderators were able to "follow test users' lead" by extemporaneously prompting for additional information or checking assumptions when indicated. Through the Blackboard Collaborate interface, moderators and test users could (via desktop computers) see, hear, show, capture, complete questionnaires, and engage in different kinds of interpretive dialogue. More importantly, the Blackboard Collaborate interface provided moderators with the opportunity to define when and how that interpretive dialogue took shape [28].

Conducting robust usability testing of a complex Internet-based intervention using Web-conferencing software that was new to test users did, however, create a cognitively demanding environment. Moderators had to manage concurrent use of Blackboard Collaborate plus the online intervention being tested, all while trying not to confuse learners or overload themselves. The real-time aspect meant that usability sessions were never predictable. The need for moderators and test users to divide their attention among auditory, textual, and visual material made high demands on limited working memory, creating at times a kind of "cognitive overload":

Especially at the start of the session, when you were trying to get everything set up and working properly for their audio [and] ...explain how the 'think aloud' process worked,...there was a lot to keep track of on the screen...One time a test user forgot to unmute and just started talking...we had no idea...There were quite a few interruptions in the first 5 to 10 minutes." [MOD 1]

Moderators also had to adapt their approaches to each test user's responses and needs. Moderators in Project 1, which involved the same group of test users in both testing cycles, noted that these challenges were greatly reduced during the second cycle of testing as everyone had more experience with Blackboard Collaborate and with the usability testing process. Moderators' perceptions were that Cycle 2 was not effortless, but certainly more efficient:

> *[It was] more relaxed...conversation was more focused...more time could be spent exploring possible solutions to problems.... [We were] less anxious about technical problems...[and] didn't feel as stressed.* [MOD 2]

One moderator noted how Blackboard Collaborate sessions meant "being ready for anything" and having to problem-solve "on the fly," although moderators generally felt they were less affected by "cognitive overload." All moderators talked about their use of the mute button that allowed them to listen without being heard. Moderators felt this "privacy" allowed them to keep the live testing space quiet and less distracting (eg, test users didn't hear them drinking water, there was less background noise to distract the test user). All moderators provided anecdotal examples where test users themselves used the mute button to attend to something happening outside the test session (eg, receiving a phone call, checking on somebody in their house, eating lunch). Moderators in Project 1 used the "private chat" function as another channel to communicate between themselves without "disrupting" the test user:

> *Private chat was helpful...we could keep each other on track for time...When users were busy working on a task, we could private message each other. The test users didn't see [that] we could check in with each other.* [MOD 1]

The ability to control and create these mini "offline" experiences in the online space meant moderators and test users could attend to other (and sometimes outside) impromptu demands. Moderators experienced fluctuating demands on their mental resources across test user sessions and testing cycles. Keeping test users on task, dealing with simultaneous tasks, and optimizing use of time during the session, required that moderators have considerable capacity to problem-solve in a complex collaborative environment.

Theme 2: Moderators Experienced Challenges in Creating and Sustaining a Sense of "Presence" and Untangling Dialogue

Moderators expressed different opinions about the quantity and nature of the social interaction, or sense of "presence," that Blackboard Collaborate supported, both across the different projects and across iterative testing cycles. Table 2 outlines the main benefits and limitations experienced by moderators in creating and sustaining different facets of presence: social presence (ie, the sense of being with others), control (ie, the sense of interacting in an environment that is responsive to you), and personal presence (ie, the sense of immediacy or "being there"). Despite the synchronous nature of the exchanges in Blackboard Collaborate that mimic the interaction possibilities in face-to-face testing, there were still challenges for moderators and test users in terms of quickly establishing rapport in a virtual environment. The ability of test users to control the flow of communication to some degree (ie, choosing whether to use the camera tool, muting the session momentarily if needed, and completing the test from any location) meant that "presence"

in the testing environment reflected test users' own choices—not a predetermined "ideal" test environment.

Table 2. Examples of benefits and limitations of using Blackboard Collaborate.

Presence factors		Specific examples
Benefits	Anonymity	Moderators or test users might be more willing to share honestly or critically if less visible/identifiable.
	Test users have a sense of control	Moderators or test users could mute the session if they wanted to limit noise.
		Moderators could employ the "private chat" feature.
		Test users have control of when and where their session was held (some test users completed testing at home or at work).
	Authentic use in context	Test users could mute the session for reasons such as: check on children, take a phone call, speak to a coworker, eat.
		By having an unstandardized testing approach, the teams were given insights into the nature of technology use in people's everyday lives and routines. This was valuable information about how the eHealth interventions being tested might also be used.
Limitations	Lack of visual cues, "personality," or human element in the virtual space	The lack of visual cues led to moderators feeling they were checking in with the test user more than necessary. If there was silence, or no movement on the screen, moderators couldn't be sure if the test user was done or just attending to another task.
	Quickly establishing rapport and relationships	At times, moderators experienced anxiety about getting technical problems solved quickly to reduce test user stress and to ensure not too much testing time was taken up by technical problems. In Project 1, the moderator was on the same campus as some of the test users and was requested to come in person to the test user's office to set up the audio prior to the test session.
	Concept of time	The technical setup took longer than anticipated, so at times moderators felt rushed for time to complete usability tasks.
		There was no "clock" tool to help provide test users or moderators with cues about how much time a task had taken.
	Surveillance	Moderators' virtual presence was constant and all-encompassing. Test users' every click was monitored and every task was recorded. Moderators felt that the testing context might have led to feelings of being surveilled, obligations to have opinions, or pressure on test users to perform as expected.

While all moderators were eager to allay test user anxieties or help overcome challenges the test users might be experiencing, they also did not "want to interject too often as the goal of the session was to identify problems" [MOD 3]. Tangled conversation (ie, speaking over each other, unintentionally interrupting) was exacerbated by technical problems with audio and video play that sometimes cut-out completely or lagged, resulting in episodes of audio speeding up in order to "catch up" or audio feedback. In Project 1, testing sessions were more moderator-led, with less time for test users to explore freely and more structured interaction between the test user and moderator. Given that usability sessions had a target time limit (eg, 90 minutes) moderators needed to manage the sessions closely. It is interesting that all test users opted out of using the Simultaneous Camera feature in Blackboard Collaborate, meaning they did not see the moderator and, therefore, had no eye contact or body language to inform their communication strategies. The physical or "personal" disconnection was noted as an important factor in Project 1 more than in Project 2. Given that Project 2 had only 1 moderator and 1 test user in each session, it may be that there was less need for explicit social feedback to manage orderly conversations even if pausing frequency was sporadically difficult to gauge:

> *...it could be pretty quiet...just you there listening and they were working through tasks.... You'd need to check in and make sure everything was working OK if they didn't say anything for a while.... I asked questions, then they asked questions...then I asked questions.* [MOD 3]

Project 2, which incorporated a largely uninterrupted, free-exploration opportunity for test users, allowed the moderator to use Blackboard Collaborate as more of a remote observation tool and less as an interactive communication tool. As Project 2 only involved test users in a single session with 1 moderator, there was less opportunity (or arguably need) to take advantage of all the advanced communication features. The moderator observed how the test user was interacting with the online intervention through a shared desktop and could answer questions verbally as or if needed. Since there was only 1 moderator and a single user in each session, the moderator found there were fewer relational dynamics to manage and fewer interaction cues to monitor (eg, who was logged in, who was speaking, who was typing in the chat box).

In Project 1, which had 2 moderators present, each moderator only had a limited amount of power to direct where the conversation went. Each moderator in that case was charged with leading a certain aspect of the test sessions, with priorities and assumptions in the mix, meaning the exchange could be pulled in any number of directions. Conversations could jump around and move away from a topic a moderator was getting

ready to talk about. In the absence of visual cues, moderators in Project 1 expected that people might "talk over each other" at times and were reluctant to interject and be seen as "disrupting" or "talking over" someone else. Not infrequently, the moderators and test users interrupted each other, but for the most part there was a comfortable back and forth.

Despite only a small number of people attending usability sessions, moderators identified conversational challenges. Moderators from both projects felt remote communication mediated by the Web-conferencing tool led to more simultaneous talking and "tangled" conversation. They noted some difficulties in managing both overenthusiastic and silent test users along with offering the right level of support during think-aloud exercises. A moderator from Project 1 provided an example:

> *Sometimes you would talk and there was a little delay...the other person would start talking and it would get confusing...sometimes [you] needed to wait and make sure they were finished talking or else you would end up talking over and interrupting each other...and some people were really chatty and it was hard to read when to get a word in...to interject and refocus.* [MOD 1]

Theme 3: Moderators Experienced Diverse Technical Demands but Were Able to Collaboratively Troubleshoot With Test Users

Technical considerations related both to the technical infrastructure as well as the technical competency of moderators and test users. Although the overall computer competency of the test users was quite high, many had never used Web conferencing before. Given that test users were both evaluating an online eHealth intervention they were unfamiliar with and using a Web-conferencing tool, technical issues arose and developing collaborative dialogue was challenging at times for moderators. Moderators found that they were not only required to make more advanced use of the interface during the session, but also were ultimately responsible for providing live troubleshooting support. Moderators described experiencing the most significant technical difficulties around ensuring high quality audio (eg, reducing audio feedback, volume, clarity), software requirements and compatibility issues, and data export (eg, file format) for further data analysis.

> *You needed backup plans.... One participant was supposed to update their JavaScript before the session but didn't.... We tried for 10 or 15 minutes and couldn't get that fixed...[so] we ended up having them switch to a different computer altogether.... We were wasting time.* [MOD 1]

Moderators also expressed a certain degree of stress in "rushing" or trying to "get through" the usability protocol tasks given more time than expected had to be spent on technical issues with some test users. All moderators described instances of being affected by what they perceived as test user stress and varying degrees of technical computer competence. Moderators in Project 1 felt that they:

> *...didn't really have time to learn and incorporate the 'bells and whistles' [of Blackboard Collaborate]...like the emoticons...[which] might have helped communication, but it would have taken time for...[test users] to learn how to use them.* [MOD 2]

All moderators tended to downplay the overall impact of these technical issues in terms of the quality of usability results. Technical inconveniences were primarily experienced during setup. Moderators also experienced interactions with test users that included laughing, light-hearted joking about technical prowess, and opportunities to empathize with test users around technical malfunctions.

Discussion

Principal Results

Moderators' experiences across both projects in this study have identified functional advantages and disadvantages of using Web-conferencing software for usability testing. Much has been written about how Web conferencing allows the moderator to "capitalize" on functionality that supports interaction and collaboration [29]. However, the need for moderators and test users to divide their attention among auditory, textual, and visual material makes high demands on limited working memory and may result in cognitive overload [30,31]. This kind of "disciplined improvisation" [32] presented challenges for the 2 eHealth projects we examined. Traditional usability techniques such as think-aloud and cognitive walk-through are not easily applied in dynamic, interruption-prone environments or with clinical populations who may have complex underlying motivational, cognitive, or physical challenges [33,34]. Test users in these contexts are evaluating a health technology that they are not familiar with via a usability testing mechanism that is also unfamiliar, which can create cognitively demanding test scenarios. Project 1 emphasized the need to consider cognitive load when developing usability testing methods. This would suggest that moderators engaged in lengthy usability sessions need advanced skills and knowledge to navigate the clinical, technical, online collaboration, and software development process aspects of the test sessions. Researchers should acknowledge how human factors not only affect the design and implementation of health interventions, but also the testing and vetting processes as well. The emergence of many new usability services and processes provide promising technical facilitation opportunities, but there is a need for more research evidence about how the nature of these virtual testing environments might be mediating or moderating results in unexpected ways.

Limitations

The advanced feature set of Blackboard Collaborate (eg, Web cams, polling, emoticons) might have helped improve collaboration, but wasn't pragmatic in short usability testing sessions like those described here, where there was limited time to learn about and develop competency in all the features. Unlike a semester-long, Web-conference-delivered course where facilitators and learners have significant time to develop proficiency and use the more advanced features of a tool, our short usability sessions required moderators and test users to quickly adapt and learn the technology. The technical and

collaborative competencies of the moderator and test users may be particularly amplified for shorter usability sessions like those in our projects.

In this case study, the moderators' challenges in creating and sustaining social presence and untangling dialogue required them to draw on their technical and interpersonal communication skills. Increased levels of interactive complexity require heightened levels of online collaborative competencies, which are supported by many online learning models [11] and may also transfer to remote synchronous usability evaluation contexts. Emerging research into self-disclosure, group norms in online communication, and use of strategies to overcome lack of nonverbal cues in computer-mediated communication [35,36,37] have potential implications for the reliability and validity of testing usability of eHealth interventions "in context" in remote online test environments. New and innovative usability platforms and services that provide access to massive pools of standardized test user data are invaluable but are not without their own limitations and bias. Researchers should not assume that any given testing strategy (ie, laboratory or remote, automated or moderated) will remove these challenges completely. This study highlights some of the human factors that shape interaction between moderators and test users in a Web-conference test environment. Given evidence in the online learning literature of the relationship between interaction and perceived effectiveness, user satisfaction, and engagement [38], it is important to better understand how interaction patterns observed in moderated remote testing might affect interpretation of usability results and ultimately influence design decisions made as a result. A more rigorous research program in the field of moderating usability testing for eHealth interventions could lead to improved training, the development of better testing tools/platforms, and more refined usability measures.

Findings suggest that moderators of usability sessions face diverse technical demands but are, if experienced with the technology, able to collaboratively troubleshoot with test users. Research suggests that some of these interaction challenges create additional stress for moderators. Moderators might infer that test users need assistance, but it can be difficult to know when to interject or offer support [39] without confounding usability test results. While some would suggest that a nonexpert moderator's failure to understand subtle features of the tool or its use might have a crippling impact on the usability session, we found that, generally, test users and moderators demonstrated considerable technical and collaborative competencies to preempt technical challenges or troubleshoot and resolve issues together during the sessions. Perhaps ironically, technical challenges seemed to create opportunities to show empathy and humanize the moderator-tester relationship—particularly in the first few minutes of the test session when rapport was just being established.

Conclusions

If current trends continue, the general population will become increasingly familiar with Web-conferencing tools through formal education [40]. Moderators will increasingly require online collaborative skills to navigate test user needs, resolve technical challenges, and accommodate "real life" events that may unexpectedly appear during the testing process. As competencies with Web conferencing increase, many of the issues highlighted in this paper might be overcome and the benefits of remote testing more easily realized.

It may be helpful to formally investigate the relationships between moderators' experiences and their personal characteristics (including previous usability experience and private theories about "good design") to help researchers understand how to best prepare moderators to support test users in virtual environments. While laboratory settings allow for more experimental control and collection of various types of data, these settings lack the realism of a field setting and deeper engagement from potential end users into hedonic quality factors that impact satisfaction [41]. "How-to" books and resources for facilitating and moderating usability sessions and training programs for usability testers are becoming more accessible, but published peer-reviewed research on moderator experiences is lacking. Insights from moderator-focused studies might be advantageous in designing test environments that put less cognitive burden on all test scenario test users. Comparative research is also needed to better understand how cognitive load and technical competence might moderate results in remote versus in-person usability testing [42]. Understanding a moderator's role in usability testing, as well as the influences and impacts their role can have, requires a sociotechnical framework that accounts for the complex interactions between human behavior and actions and the tools and technologies in the environment [43].

Acknowledgments

This project is funded by an operating grant from the Canadian Institutes of Health Research (CIHR; MOP119531). Dr Newton holds a CIHR New Investigator Award.

Authors' Contributions

ASN conceived of the study, acquired funding, participated in its design and coordination, and helped to draft the manuscript. LW participated in the drafting of the study protocol and drafted the manuscript. PB participated in data collection/analysis and made significant contributions to the manuscript. HF, LC, and AH participated in data collection/analysis and helped edit the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]
2. Hertzum M, Molich R, Jacobsen N. What you get is what you see: revisiting the evaluator fffect in usability tests. Behaviour & Information Technology 2014;33(2):143-161. [doi: 10.1080/0144929X.2013.783114]
3. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Medical Inform Assoc 2012;19(3):413-422.
4. Dumas J, Loring B. Moderating usability tests: principles and practice for interacting. Amsterdam: Morgan Kaufmann/Elsevier; 2008.
5. Molich M, Wilson C. Tips and tricks for avoiding common problems in usability test facilitation. In CHI “08: CHI “08 extended abstracts on Human factors in computing systems. New York, NY: ACM; pp 2008:2379-2382. [doi: 10.1145/1358628.1358689]
6. Brock D, Kim S, Palmer O, Gallagher T, Holmboe E. Usability testing for the rest of us: the application of discount usability principles in the development of an online communications assessment application. Teach Learn Med 2013;25(1):89-96. [doi: 10.1080/10401334.2012.741547] [Medline: 23330901]
7. Chang HS, Tsai TH, Chang YC, Chang YM. Touch panel usability of elderly and children. Computers in Human Behavior 2014 Aug;37:258-269. [doi: 10.1016/j.chb.2014.04.050]
8. Andreasen MS, Neilsen HV, Schroder SO, Stage J. What happened to remote usability testing? An empirical study of three methods. In: Human Factors in Computing Systems Conference. NY: ACM; 2007:1405-1414.
9. Brush AJ, Ames M, Davis J. A comparison of synchronous remote local usability studies for an expert interface. In: CHI 04 extended abstracts on Human factors in computing systems. 2004 Presented at: Human Factors in Computing; April 24-29; Vienna, Austria p. 1179-1182. [doi: 10.1145/985921.986018]
10. Thomsett-Scott B. Web site usability with remote users. Journal of Library Administration (3/4) 2006;45:517-547.
11. Dixon E. Building a model for online distance courses through social media and networks. International Journal of Virtual and Personal Learning Environments 2012;3(3):81-94.
12. Andrzejczak C, Liu D. The effect of testing location on usability testing performance, participant stress levels, and subjective testing experience. Journal of Systems and Software 2010 Jul;83(7):1258-1266. [doi: 10.1016/j.jss.2010.01.052]
13. Lesaigle E, Biers D. Effect of Type of Information on Real Time Usability Evaluation: Implications for Remote Usability Testing. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2000 Jul 01 Presented at: the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society; Oct 8-12; Santa Monica, CA p. 585-588. [doi: 10.1177/154193120004403710]
14. Cornelius S. Facilitating in a demanding environment: Experiences of teaching in virtual classrooms using web conferencing. Br J Educ Technol 2013 Feb 07;45(2):260-271. [doi: 10.1111/bjet.12016]
15. Kear K, Chetwynd F, Williams J, Donelan H. Web conferencing for synchronous online tutorials: Perspectives of tutors using a new medium. Computers & Education 2012 Apr;58(3):953-963. [doi: 10.1016/j.compedu.2011.10.015]
16. Trivedi M. Role Of Context In Usability Evaluations: A Review. ACIJ 2012 Mar 31;3(2):69-78. [doi: 10.5121/acij.2012.3208]
17. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]
18. van den Haak MJ, de Jong MDT. Analyzing the interaction between facilitator and participants in two variants of the think-aloud method. In: In Professional Communication Conference Proceedings. 2005 Presented at: Professional Communication Conference; Jul 10-13; Limerick, Ireland p. 323-327. [doi: 10.1109/IPCC.2005.1494192]
19. Wozney L, McGrath P, Newton A, Huguet A, Franklin M, Perri K, et al. Usability, learnability and performance evaluation of IRIS (Intelligent Research and Intervention Software): a delivery platform for eHealth interventions. Health Informatics J 2015 Jun 23. [doi: 10.1177/1460458215586803]
20. Yin R. Case study research: Design and methods. Thousand Oaks, Calif: Sage; 2003.
21. Baxter P, Jack S. Qualitative case study methodology: study design and implementation for novice researchers. Qualitative Report 2008;13(4):544-559.
22. Blackboard Collaborative Online Teaching Platforms | Blackboard Collaborate. Blackboard Collaborate URL: http://www.blackboard.com/online-collaborative-learning/index.aspx [accessed 2016-01-04] [WebCite Cache ID 6eHvJdYmI]
23. Koelsch LE. Reconceptualizing the member check interview. Int J Qual Methods 2013;12:168-179.
24. Willig C. Quality in qualitative research. In C. Willig (Ed.), Introducing qualitative research in psychology (2nd ed., pp. 149-161). Maidenhead: Open University Press; 2008.

25. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011;11:100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
26. Silverman D. Interpreting qualitative data: methods for analysing talk, text and interaction. Thousand Oaks, CA: Sage Publications; 1995:437-438.
27. Miles M, Huberman A. Qualitative data analysis: an expanded sourcebook. Thousand Oaks, CA: Sage Publications; 1994.
28. Følstad A, Hornbæk K. Work-domain knowledge in usability evaluation: Experiences with Cooperative Usability Testing. Journal of Systems and Software 2010 Nov;83(11):2019-2030. [doi: 10.1016/j.jss.2010.02.026]
29. McBrien J, Cheng R, Jones P. Virtual spaces: employing a synchronous online classroom to facilitate student engagement in online learning. International Review of Research in Open and Distance Learning 2009;10(3):1-17.
30. Mayer R, Heiser J, Lonn S. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. Journal of Educational Psychology 2001;93(1):187-198. [doi: 10.1037//0022-0663.93.1.187]
31. Kirschner F, Paas F, Kirschner P. Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. Computers in Human Behavior 2009 Mar;25(2):306-314. [doi: 10.1016/j.chb.2008.12.008]
32. Sawyer R. Creative teaching: collaborative discussion as disciplined improvisation. Educational Researcher 2004;33(2):12-20.
33. Kellar M, Hawkey K, Inkpen K, Watters C. Challenges of Capturing Natural Web-Based User Behaviors. Int. J. of Human-Comp. Interaction 2008 May;24(4):385-409. [doi: 10.1080/10447310801973739]
34. Bastien J M Christian. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010 Apr;79(4):e18-e23. [doi: 10.1016/j.ijmedinf.2008.12.004] [Medline: 19345139]
35. Jiang LC, Bazarova NN, Hancock JT. The disclosure-intimacy link in computer-mediated communication: an attributional extension of the hyperpersonal model. Human Communication Research 2011;37(1):58-77. [doi: 10.1111/j.1468-2958.2010.01393.x]
36. Antheunis M, Schouten A, Valkenburg P, Peter J. Interactive Uncertainty Reduction Strategies and Verbal Affection in Computer-Mediated Communication. Communication Research 2011 Jun 01;39(6):757-780. [doi: 10.1177/0093650211410420]
37. Yun G, Park S, Holody K, Yoon K, Xie S. Selective Moderation, Selective Responding, and Balkanization of the Blogosphere: A Field Experiment. Media Psychology 2013 Jul;16(3):295-317. [doi: 10.1080/15213269.2012.759462]
38. Tsai C, Shen P, Chiang Y. Research trends in meaningful learning research on e-learning and online education environments: A review of studies published in SSCI-indexed journals from 2003 to 2012. Br J Educ Technol 2013 Oct 17;44(6):E179-E184. [doi: 10.1111/bjet.12035]
39. Nelson E, Stavrou A. Advantages and Disadvantages of Remote Asynchronous Usability Testing Using Amazon Mechanical Turk. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2011 Sep 01 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting ;55(1); 2011; Las Vegas p. 1080-1084. [doi: 10.1177/1071181311551226]
40. Lê T, Lê Q. Technologies for enhancing pedagogy, engagement and empowerment in education: creating learningâ€ friendly environments. Hershey, Pa: IGI Global; 2012.
41. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055]
42. Weinel M, Bannert M, Zumbach J, Hoppe H, Malzahn N. A closer look on social presence as a causing factor in computer-mediated collaboration. Computers in Human Behavior 2011 Jan;27(1):513-521. [doi: 10.1016/j.chb.2010.09.020]
43. Patel VL, Kannampallil TG. Human factors and health information technology: current challenges and future directions. Yearb Med Inform 2014;9:58-66 [FREE Full text] [doi: 10.15265/IY-2014-0005] [Medline: 25123724]

Abbreviations

CIHR: Canadian Institutes of Health Research
FASD: fetal alcohol spectrum disorder
IRIS: Intelligent Research and Intervention Software

Edited by G Eysenbach; submitted 01.05.15; peer-reviewed by T Herlihey; comments to author 26.07.15; revised version received 21.09.15; accepted 07.10.15; published 03.02.16.

Please cite as:
Wozney LM, Baxter P, Fast H, Cleghorn L, Hundert AS, Newton AS
Sociotechnical Human Factors Involved in Remote Online Usability Testing of Two eHealth Interventions
JMIR Human Factors 2016;3(1):e6
URL: http://humanfactors.jmir.org/2016/1/e6/
doi:10.2196/humanfactors.4602
PMID:27026291

Original Paper

The Usability of Diabetes MAP: A Web-delivered Intervention for Improving Medication Adherence

Lyndsay A Nelson[1,2], PhD; Magaela C Bethune[3], MS, MPA; Andrea E Lagotte[1,2], MA; Chandra Y Osborn[1,2,4,5], PhD, MPH

[1]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Center for Health Behavior and Health Education, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Department of Human & Organizational Development, Vanderbilt University, Nashville, TN, United States

[4]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[5]Center for Diabetes Translational Research, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Chandra Y Osborn, PhD, MPH
Center for Health Behavior and Health Education
Vanderbilt University Medical Center
2525 West End Ave, Suite 370
Nashville, TN, 37203
United States
Phone: 1 615 936 8468
Fax: 1 615 875 2201
Email: chandra.osborn@vanderbilt.edu

Abstract

Background: Web-delivered interventions are a feasible approach to health promotion. However, if a website is poorly designed, difficult to navigate, and has technical bugs, it will not be used as intended. Usability testing prior to evaluating a website's benefits can identify barriers to user engagement and maximize future use.

Objective: We developed a Web-delivered intervention called Diabetes Medication Adherence Promotion (Diabetes MAP) and used a mixed-methods approach to test its usability prior to evaluating its efficacy on medication adherence and glycemic control in a randomized controlled trial.

Methods: We recruited English-speaking adults with type 2 diabetes mellitus (T2DM) from an academic medical center who were prescribed diabetes medications. A trained research assistant administered a baseline survey, collected medical record information, and instructed participants on how to access Diabetes MAP. Participants were asked to use the site independently for 2 weeks and to provide survey and/or focus group feedback on their experience. We analyzed survey data descriptively and qualitative data thematically to identify participants' favorable and unfavorable experiences, characterize usability concerns, and solicit recommendations for improving Diabetes MAP.

Results: Enrolled participants (N=32) were an average of 51.7 ± 11.8 years old, 66% (21/32) female, 60% (19/32) non-Hispanic White, 88% (28/32) had more than 12 years of education, half had household incomes over $50,000, and 78% (25/32) were privately insured. Average duration of diagnosed diabetes was 7.8 ± 6.3 years, average A1c was 7.4 ± 2.0, and 38% (12/32) were prescribed insulin. Of enrolled participants, 91% (29/32) provided survey and/or focus group feedback about Diabetes MAP. On the survey, participants agreed website information was clear and easy to understand, but in focus groups they reported navigational challenges and difficulty overcoming user errors (eg, entering data in an unspecified format). Participants also reported difficulty accessing the site and, once accessed, using all of its features. Participants recommended improving the site's user interface to facilitate quick, efficient access to all features and content.

Conclusions: Adults with T2DM rated the Diabetes MAP website favorably on surveys, but focus groups gave more in-depth feedback on the user experience (eg, difficulty accessing the site, maximizing all of the site's features and content, and recovering from errors). Appropriate usability testing methods ensure Web-delivered interventions work as intended and any benefits are not diminished by usability challenges.

(JMIR Human Factors 2016;3(1):e13) doi:10.2196/humanfactors.5177

KEYWORDS

Website; Usability Testing; Type 2 Diabetes Mellitus; Medication Adherence; Intervention

Introduction

Among adults with type 2 diabetes (T2DM), approximately 1 in 3 do not take their medications as prescribed [1], and nonadherence is associated with suboptimal glycemic control [2], hospitalizations [3,4], and pre-mature death [4,5]. Very few interventions improve medication adherence, and among those that do, effects are generally small [6]. Moreover, most efficacious interventions have been delivered face-to-face, making them more labor-intensive and less feasible in busy clinic settings [7]. An estimated 84% of adults in the United States use the Internet [8], so the automated nature of Web-delivered interventions makes them a more feasible alternative to face-to-face approaches [9].

Web-delivered interventions have mixed effects on health behaviors [10-12] and varied effects on glycemic control [13,14]. Wide variability in both the time spent using websites and how they are used may explain their mixed effects on health behaviors and outcomes [15]. Website engagement varies widely between studies [16,17], and more engagement is often associated with greater improvement in outcomes [18,19]. A fundamental determinant of website engagement is a website's usability [20], or how easy a user interface is to use.

The evaluation of a website's usability is necessary before testing its potential efficacy on health behaviors and outcomes [21]. Website usability is the extent to which users can effectively, efficiently, and satisfactorily interact with a website [22]. Six factors determine a site's usability, including (1) an intuitive design (ie, the site is easy to understand and navigate), (2) its ease of learning (ie, how quickly a user can learn basic site tasks), (3) its efficiency of use (ie, how quickly a user can complete site tasks), (4) its error frequency and severity (ie, how often users make errors, the seriousness of the errors, and how users recover from errors), (5) its memorability (ie, how well a user can remember the site to use it effectively in the future), and (6) its subjective satisfaction (ie, how much the user enjoys using the site) [22]. Usability testing focuses on measuring a website's capacity to excel in each of these 6 areas.

Usability testing ensures a Web-delivered intervention works as intended, so the target audience uses it to the degree needed to reap its potential benefits [15]. Usability testing studies often employ quantitative surveys, but a qualitative approach can reveal more usability problems and concerns than surveys alone [21,23]. A mixed-methods approach includes both and provides a comprehensive assessment of website usability. Therefore, we used a mixed-methods approach focused on the 6 usability areas [22] to: (1) identify the favorable and unfavorable aspects of the Diabetes Medication Adherence Promotion (Diabetes MAP) website, including its usability challenges, and (2) solicit ideas for improving the site's usability prior to evaluating its impact on medication adherence and glycemic control in a randomized controlled trial.

Methods

Diabetes MAP Intervention

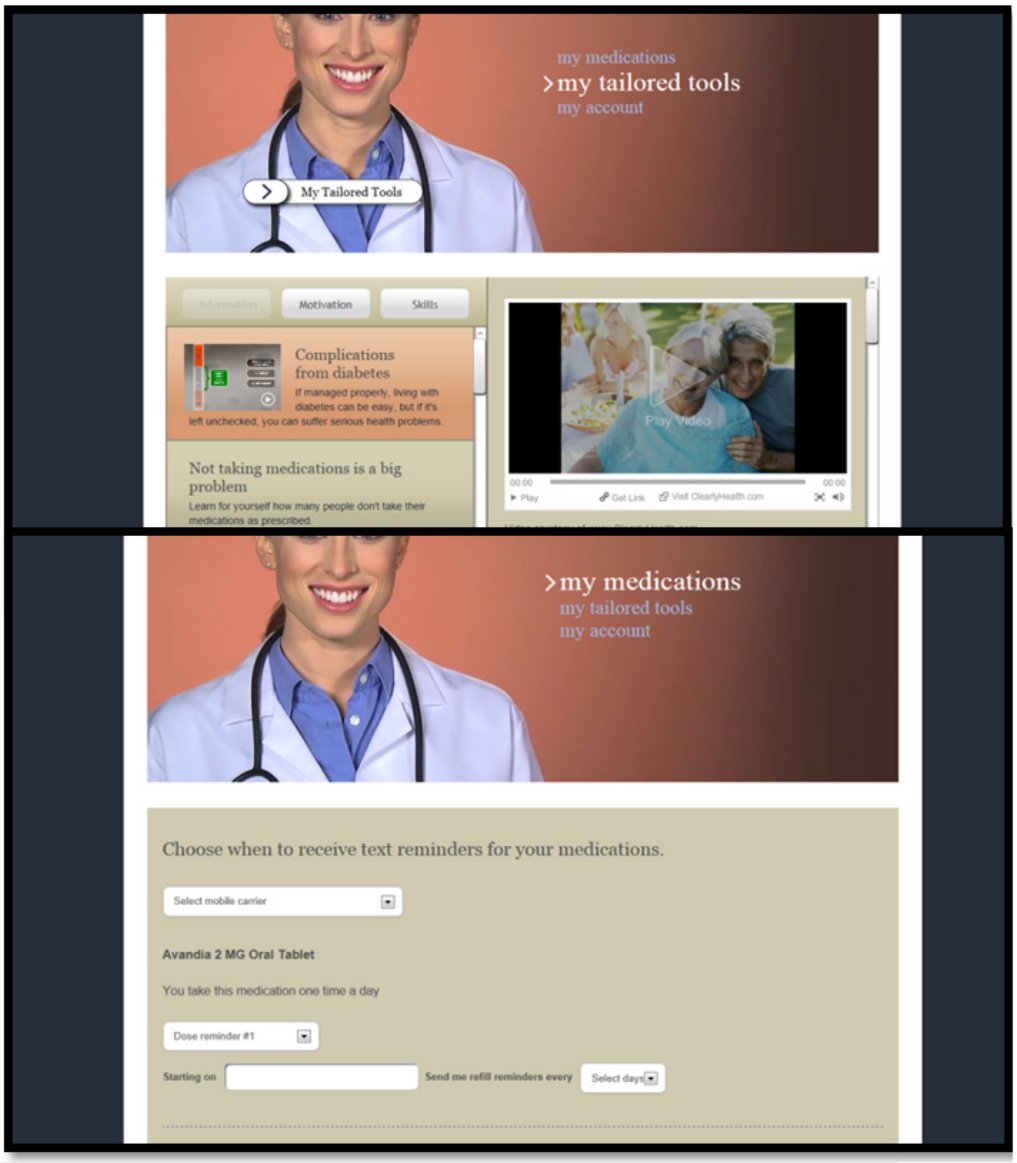
Diabetes MAP is a self-guided, Web-delivered intervention designed to promote medication adherence among patients with T2DM. Diabetes MAP's content is grounded in the Information-Motivation-Behavioral skills (IMB) model of medication adherence [24,25]. Studies in diabetes [24,25] and other chronic disease contexts (eg, HIV) [26] suggest a patient's medication adherence depends on his or her adherence-related information, motivation to take medications, and adherence-related behavioral skills. Therefore, Diabetes MAP's intervention content addresses user-specific barriers to adherence in each of these domains. Upon registering for an account and logging in to the Diabetes MAP site, users are asked to create a medication list by searching for RxNorm-generated medications. Next, they are asked a series of questions to assess their medication adherence-related information, motivation, and behavioral skills barriers. Entered medications and responses to these questions populate a separate page titled, My Tailored Tools (Figure 1, Top Panel). This page responds to a user's early inputs (eg, medications entered, barriers to adherence) with a toolbox of tailored regimen-specific and IMB-model based intervention content.

My Tailored Tools houses 30 educational videos and 11 pieces of static content to enhance user-specific adherence-related information, motivation, and behavioral skills. Informational content is both medication class-specific (eg, a video on the key facts about metformin and how it works in the body for patients prescribed metformin, a video of how insulin works in the body for patients prescribed insulin) and conveys the importance of adherence for glycemic control and preventing complications (eg, a video showing the complications that can occur from not taking medications as prescribed). Motivational content is intended to enhance patients' personal and social motivation for adherence (eg, a video on how to overcome one's fear of needles, static content presenting strategies for soliciting social support for adherence). Finally, behavioral skills content provides practical "how to" advice to ensure successful adherence (eg, a video with step-by-step instruction on how to inject insulin, a video on how to store insulin).

The Diabetes MAP website has additional capabilities. Its features (ie, functionality built into the site to enhance the user experience) allow users to perform various tasks (ie, clearly defined assignments to complete within a website). For example, upon creating a medication list, users can print and email this list, learn about each medication listed, set up medication dosing and refill reminders sent as text messages to their mobile phone (Figure 1, Bottom Panel), and connect to a patient portal account (ie, My Health at Vanderbilt) to communicate with healthcare providers about medication side effects and prescription reauthorizations via secure messaging. As noted above, users can also complete an IMB model-based barriers-to-adherence assessment and view user-specific educational videos and

content to address users' IMB model-based barriers. Finally, the website includes navigational videos explaining the site's features and giving instructions on how to complete tasks. Diabetes MAP was not designed for a specific user, but we made design choices to account for potential literacy, visual, and auditory limitations of all users. Such choices include presenting simplified language in large font, the option to watch and/or listen to videos or read video scripts, and a full-screen option to improve video visibility.

Figure 1. Top Panel: Diabetes MAP screenshot of the My Tailored Tools page presenting videos and content to address a user's barriers to adherence. Bottom Panel: Diabetes MAP screenshot of the page where a user can set up text message medication reminders.



Participants and Recruitment

To test the usability of the Diabetes MAP site, we recruited English-speaking adults from an academic medical center who were diagnosed with T2DM, prescribed diabetes medications, and had Internet access to participate in a mixed-methods study. Recruitment strategies included advertisements about the study, referrals from healthcare providers, medical center listserv announcements, and approaching patients waiting in an adult diabetes specialty clinic or adult primary care clinic. The total number of participants enrolled (N=32) reached the target enrollment for qualitative (at least 5) and quantitative (at least 20) usability testing [27,28]. The Institutional Review Board at Vanderbilt University Medical Center approved all study procedures prior to participant enrollment.

Procedures

A trained research assistant (RA) scheduled interested and eligible participants to meet individually in a private room at the medical center. The RA administered (1) informed consent, (2) a survey by reading survey items and response options out loud or by distributing one available in paper-pencil format or electronic format via Research Data Capture (REDCap™) [29] that could be completed independently, and (3) a 1-page instruction guide on how to locate and access the Diabetes MAP website, which each participant was asked to independently use

for 2 weeks. With permission, the RA also reviewed each participant's medical record to collect clinical data.

We used Mouseflow ApS™ to measure participants' use of Diabetes MAP. After 2 weeks, the RA invited participants to provide feedback on the site's usability by completing a 20-minute survey and attending a 60-minute focus group session. The survey could be completed in REDCap™ via an email link or in-person immediately before the focus group session. A trained focus group facilitator used semistructured a priori questions to assess participants' experiences with Diabetes MAP, demonstrate the site on a projector screen, and elicit responses and impressions of the site. This method is consistent with the pluralistic walkthrough approach to usability testing that involves stepping through a system with users to understand their perceptions of and experiences with a system [30,31]. The pluralistic walkthrough approach reveals users' uncertainty with a system's features and tasks better than other usability methods [31]. We asked participants if they had challenges with using the site, their most and least favorite aspects of the site, their perceived benefits of using the site, and any recommendations they had for improving it. All sessions were audio-recorded. Recordings were transcribed verbatim and de-identified prior to analyses. We compensated participants up to $155 for completing a survey at enrollment ($25), using Diabetes MAP ($8 per hour, up to 10 hours), completing the follow-up survey ($15), and participating in a focus group ($35).

Measures

Demographic Characteristics

Demographic information included participants' age, gender, race/ethnicity, years of education, annual household income, and health insurance status. We asked participants whether they owned a mobile phone and used text messaging with their phone to better understand participants' reasons for setting up or not setting up medication dosing and refill text message reminders in Diabetes MAP.

Clinical Characteristics

Participants self-reported duration of diagnosed diabetes in years and months, and the number and type of diabetes medications prescribed, including insulin. The RA reviewed each participant's medical record to confirm a T2DM diagnosis and the quantity and type of prescribed medications, and to collect participants' most recent glycated hemoglobin A1c test result to characterize the sample's glycemic control.

Website Usage

We assessed participants' website usage with data logged by Mouseflow ApS™. Specifically, we assessed the total number of days users initiated a session by logging into Diabetes MAP, the total number of minutes users were logged into Diabetes MAP (ie, from the time they created an account until the study was over), and the average number of minutes logged in per days logged in. We also captured whether users set up text message reminders to take their medications or refill prescriptions.

Usability

We assessed Diabetes MAP's usability with 10 items adapted from the Computer System Usability Questionnaire (CSUQ) [32]. Because the CSUQ assesses the subjective usability of a general computer system, we adapted its items to specify the subjective usability of Diabetes MAP. Example items include: "Overall, it was easy to learn to use Diabetes MAP" and "When I make a mistake in Diabetes MAP, I recover easily and quickly." Response options ranged from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating more favorable usability ratings.

Data Analyses

We used SPSS version 21.0 to summarize quantitative data using means and standard deviations (SD), or frequencies and percentages as appropriate. We used selective coding [33] to identify focus group comments and conversations [34] addressing the 6 areas of website usability: (1) intuitive design, (2) ease of learning, (3) efficiency of use, (4) error frequency and severity, (5) memorability, and (6) subjective satisfaction [22]. First, we read focus group transcripts in their entirety, highlighting participant comments related to opinions about, experiences with, and suggestions for Diabetes MAP. Next, we integrated similar comments into categories. After an iterative process of integration and refinement, we mapped categories of responses onto each usability area. Units of analysis included single participant comments and multi-participant conversations reporting similar or different experiences with the site and suggestions for improving the Diabetes MAP user experience (eg, strategic placement of instructions, features and Web content).

Results

The sample (N=32) was on average 51.7 ± 11.8 years of age. Most were female (66%, 21/32), non-Hispanic White (NHW; 60%, 19/32), had at least some college education (88%, 28/32), and were privately insured (78%, 25/32); half had incomes above $50,000 (Table 1). The average HbA1c was 7.4% ± 2.0%, and 38% (12/32) were on insulin. Most participants (91%, 29/32) provided feedback about Diabetes MAP via survey and/or focus group participation (up to 5 participants per group). The characteristics of our sample reflect the characteristics of the academic medical center patient population in which they were recruited from [35]. The medical center patient population is predominately NHW (74%), well-educated (ie, over 90% having education beyond a high school degree), with relatively high incomes, and private insurance (93%) [35].

Table 1. Participant characteristics (N=32).

Characteristic		Mean ± SD or n (%)	Range
Age, y		51.7 ± 11.8	26.7-73.4
Female		21 (66)	
Race/ethnicity	White (non-Hispanic)	19 (60)	
	Black (non-Hispanic)	8 (25)	
	Hispanic	3 (9)	
	Asian	2 (6)	
Education, y		16.3 ± 2.8	12.0-24.0
Annual household income	Less than $14,999	3 (9)	
	$15,000 to $24,999	4 (13)	
	$25,000 to $49,999	9 (28)	
	$50,000 to $74,999	7 (22)	
	$75,000 or more	9 (28)	
Insurance status	Private insurance	25 (78)	
	TennCare/Medicare	6 (19)	
	No insurance	1 (3)	
Own a mobile phone		32 (100)	
Text message with phone		26 (81.3)	
Diabetes duration, y		7.8 ± 6.3	0.0-20.0
Number of diabetes medications		1.8 ± 0.8	1.0-4.0
Prescribed insulin		12 (38)	
A1c (n=31)[a]		7.4 ± 2.0	4.9-15.8

[a]One participant did not have an A1c test result in the medical record at the time of data collection.

Among all participants enrolled in the study, the average number of days users logged into the site was 4.2 ± 4.2 days during the 2-week period. The average number of hours logged into the site was 4.3 ± 4.8 hours, and the average time logged in per days logged in was 56.6 ± 47.2 minutes. Five participants (16%) set up text message reminders to take their medications and 4 participants (13%) set up text message reminders to refill their prescriptions.

Quantitative Feedback

On the survey, participants rated Diabetes MAP's usability above average (ie, scores of >3 on a 5-point scale) on each of the 10 items (Table 2). The total usability rating averaged across all items was 3.86 ± 0.90. We also mapped each survey item onto the usability area it most closely reflected. As shown in Table 2, participants rated the understandability of the site's information (ie, intuitive design), the clarity of the site's information (ie, intuitive design), and the pleasantness of interacting with the site (ie, subjective satisfaction) the most favorably. In contrast, participants rated the ease and quickness of recovering from mistakes (ie, error frequency and severity), the effectiveness of the site's information in helping users complete tasks (ie, efficiency of use), and the ease of navigating the site (ie, intuitive design) the least favorably (Table 2).

Qualitative Feedback: Usability Areas

Across 9 focus groups, participants shared experiences using Diabetes MAP, including concerns about its usability and recommendations for improvement. Of the 24 unique usability concerns reported, 14 concerns were mentioned in the first focus group and another 6 concerns were mentioned in the second focus group. By the fourth focus group, 95% of all unique usability concerns had been reported. Generally, participants' experiences with Diabetes MAP were similar across age, gender, race/ethnicity, education, and income. Table 3 presents each unique concern organized by usability area [22] and the number of focus groups it was mentioned in. The 6 areas of website usability [22] provide a framework for examining participants' usability concerns and recommendations for improvement.

Table 2. Survey items assessing Diabetes MAP's usability, ranked most to least favorably.

	Related Usability Area	Respondents[a,b], n	Mean ± SD
The information provided in Diabetes MAP is easy to understand.	Intuitive Design	28	4.3 ± 0.7
The information (such as help videos, on-screen messages, etc.) provided in Diabetes MAP is clear.	Intuitive Design	27	4.2 ± 0.8
My user interaction(s) with Diabetes MAP are pleasant.	Subjective Satisfaction	27	4.0 ± 0.9
Overall, it was easy to learn to use Diabetes MAP.	Ease of Learning	28	3.9 ± 1.0
Overall, I feel comfortable using Diabetes MAP.	Subjective Satisfaction	28	3.9 ± 1.0
It is easy to find the tools and information that I need.	Intuitive Design	27	3.8 ± 0.9
The organization of information in Diabetes MAP is clear.	Intuitive Design	27	3.8 ± 0.9
It is easy to navigate the Diabetes MAP website.	Intuitive Design	29	3.6 ± 0.9
The information provided in Diabetes MAP is effective in helping me complete tasks on the website.	Efficiency of Use	27	3.6 ± 0.9
When I make a mistake in Diabetes MAP, I recover easily and quickly.	Error Frequency and Severity	26	3.5 ± 0.9

[a]Number of participants providing a response for each item on a scale of 1 (strongly disagree) to 5 (strongly agree).

[b]Some participants indicated items were "Not Applicable" to their experience.

Table 3. Participants' concerns with Diabetes MAP by usability area.

Usability Areas	Concern	Number of Focus Groups Reporting Concern
Intuitive Design	The site's layout and content placement was confusing	4
	Unnecessary scrolling required to access site features and tasks	3
	Difficult to explore the site using the navigation menu	3
	Unclear how entering information into site tailored the user experience	1
	Unclear how to minimize navigational videos	1
	Location of navigational videos was confusing	1
Ease of learning	Instructions for accessing and using the site were unclear	4
	Directions for use within the site were unclear	3
	Navigational videos did not help with accessing features/completing tasks	3
	Navigational videos dysfunctional	1
Efficiency of Use	Unable to save progress with completing tasks	3
	Website pages took a long time to load	3
	Difficult to select time-zone using worldwide map	2
	Automatically logged out of site if stopping use for 20 minutes	2
	Website not compatible with other digital devices (eg, iPads®)	2
	Difficult to scroll through different site windows	1
Error Frequency and Severity	Website not compatible with different browsers	7
	Error messages encountered while trying to log in	6
	Difficult to search for medication names in medication list	3
	Difficult to search for medication doses in medication list	3
	Technical support was required to use website	1
Memorability	The site's purpose was unclear	4
	Website URL was confusing and made accessing the site difficult	3
Subjective Satisfaction	The site had a non-user-friendly interface	2

Intuitive Design

An intuitively designed website is easy to navigate and understand [22]. When users understand a site's layout and purpose, they can effortlessly explore it. The focus group facilitator demonstrated how all of Diabetes MAP's features and tasks were intended to work, but some participants did not fully understand the site when they used it independently. For example, some participants said Diabetes MAP's navigational videos were pleasant and helpful, but others said these videos distracted them from engaging with the most important aspects of the site. Some were unaware navigational videos were even available.

> *I didn't even realize there was a video connected to [the image] until you pointed out that arrowhead.* [62-year-old NHW male]

In another instance, participants were unclear how information they entered into the site affected their user experience. For example, data entered at account creation (eg, entering one's time zone and mobile phone number) impacts functionality elsewhere on the site (eg, receiving text message medication reminders in the appropriate time zone), and data entered into the IMB model-based barriers-to-adherence assessment impacts what videos and content are available for viewing in a user's My Tailored Tools section of the site. As a result, some participants did not access or use certain parts of the site.

It was also common for participants to miss out on site features and functions entirely (eg, the option to print one's medication list or set up text message reminders) because they were unable to locate them.

> *I am really frustrated because I would have loved [text message reminders]. I'm serious. Where was it?* [55-year-old African American/Black female]

Related to this issue were concerns with navigating between different types of content in Diabetes MAP. It was common for participants to describe difficulty reading task instructions, viewing educational videos, and using features on a single webpage. In 1 focus group, participants commiserated with one participant who said she could have used more assistance with exploring the site:

> *I [would have liked] more instructions to help navigate [the medication list] and clearly access the site.* [26-year-old Hispanic female]

Unintuitive design issues such as this one made it difficult for participants to successfully use and fully engage with Diabetes MAP.

Ease of Learning

Ease of learning refers to how fast new users of a website can learn and accomplish basic site tasks [22]. Focus group participants reported barriers to learning how to use Diabetes MAP, noting the site lacked clear, comprehensive instructions on how to perform certain tasks. When participants lacked the necessary information to accomplish basic tasks, they became frustrated.

> *[I] wasted a lot of time…. It had dragged on for 2 or 3 days when I could have actually been using [the site] and I had to contact you, which I didn't really want to have to do. It was frustrating, to say the least, and I just felt like, "What's wrong with me? What's wrong with my computer?"* [55-year-old African American/Black female]

Other participants voiced confusion without frustration such as this participant who said the directions to enter one's medications were confusing.

> *It wasn't a huge challenge, but in the medication list, it didn't specify if it wanted you to put in just your diabetes medicine or other medicines, so I put in all my medicines.... It would have been nice if it was more specific.* [27-year-old NHW female]

In the most extreme cases, some participants said the navigational videos did not help them, particularly when videos did not work. When asked about these videos, members of 1 focus group were united in their unsuccessful experience.

> *I never could get [the video] to play.* [35-year-old NHW female]
>
> *And I couldn't either, and…I thought maybe it was my computer, but it wasn't, it was [the videos] I guess.* [47-year-old, African American/Black female]

Efficiency of Use

Efficiency of use refers to how quickly a user can complete website tasks [22]. Some participants reported difficulty completing tasks in Diabetes MAP in a timely and efficient way. This was in part due to variability in website loading times on certain devices.

> *I felt like it was a little heavy to start with… iPads® can open it, but [it] needs a lot of time, even though I have high speed [Internet]. When you open it [on the] iPad®, you can't get some clips unless you are [using] a desktop or laptop.* [35-year-old NHW male]

Other participants speculated loading delays were due to the size and volume of videos being streamed.

> *[The navigational and educational] videos take a lot of feed. It takes forever to load, and when you click [one], it doesn't immediately work.* [32-year-old NHW female]

In some instances, participants were unable to save their progress on a task to revisit it and complete it later. The site also logs users out who are logged in, but who do not use the site for 20 minutes, which resulted in several participants losing task progress for partial completion.

> *If you do half of it, and you try to do something else, and the computer freezes or logs you out, you have to start all over again. Is there a possible way—I'm sure there is—to save it and come back to it to finish it?* [55-year-old NHW male]

Finally, some participants felt certain tasks were overly complicated and time-consuming. For example, the site asks users to select their time zone on a worldwide map instead of from a more efficient drop-down menu.

> *When it got to the time zone, and that little map came up…I was thinking if they had a drop down for [it]—it would probably be easier—instead of a map.* [47-year-old African American/Black female]

Error Frequency and Severity

Error frequency and severity refers to how often users make errors, the seriousness of the errors, and how users recover from errors [22]. Many participants encountered Web browser issues, made mistakes during task completion, and received error messages they did not understand or could not overcome. User errors began when first attempting to access the site. Despite receiving written instructions on the Web browser requirements for accessing Diabetes MAP, participants in 7 of 9 focus groups reported browser-related problems, and subsequent error messages. In some cases, this led participants to stop trying to access the site altogether.

> *It was really weird. I have multiple browsers of Internet Explorer®, and I kept trying to change them, thinking maybe I'm just not using the right compatibility thing. Finally, I was like, "OK, I'm just not going to look at this [website].*" [26-year-old Hispanic female]

Errors while creating and logging in to user accounts were mentioned in 6 out of 9 focus groups. To create an account, users are required to enter personal information using several entry methods including text fields and drop-down menus. If information is entered incorrectly, users receive error messages preventing further access. The recovery time needed to overcome these errors varied between participants. When participants were unable to access Diabetes MAP, some enlisted professional and nonprofessional technical support.

> *I just happened to know this computer guy who was coming in my [office] to do some other work and I asked him … I said, "Can you get this website up?" It took him a while, and this is all this man does is IT work.* [55-year-old NHW female]

Other participants reached out to study personnel who answered questions and provided remote assistance consistent with the written instructions participants were provided on how to access the site. Participants who encountered errors and did not seek assistance reported frustration and wasted time, causing some to give up using the site altogether.

Memorability

A website with memorability is one users can remember well enough to use it effectively in the future [22]. The memorability of Diabetes MAP was primarily hindered by its confusing URL. The Web address was lengthy, unintuitive, and included different types of punctuation and acronyms. Several participants mentioned difficulty with accessing the site specifically because of the URL. It was common for participants who forgot or mistyped the URL to search for the words "Diabetes MAP" in a search engine or in the search bar on the medical center's homepage. These troubleshooting techniques led users to incorrect websites and information. For example, participants who searched "Diabetes MAP" within search engines were often misdirected, leading some to online geographic maps of diabetes treatment facilities rather than the intended website intervention. Similar issues occurred when searching for the site on the medical center's homepage, as described by this participant:

> *I searched for Diabetes MAP on the medical center's site and got directions for how to get to the Diabetes Center.* [55-year-old NHW male]

Despite being both told about the website's intent and receiving an instructional handout with this information, participants in nearly half of the focus group sessions felt the website's purpose was confusing. In some cases participants forgot Diabetes MAP's purpose altogether, which led to using the website in unintended and ineffective ways.

> *I didn't even realize it was just for taking medications until we came to this focus group.* [55-year-old NHW female]

Subjective Satisfaction

Subjective satisfaction is determined by how much the user enjoys using the website [22]. While discussing participants' overall experience with using Diabetes MAP, a few participants mentioned concerns with the site's user interface. These participants said Diabetes MAP was difficult to operate and understand, and therefore unenjoyable to use. One participant expressed his frustration with Diabetes MAP's user interface:

> *I'm not an IT person, but I'm a supervisor in my department, and I do not have a problem [with computers].... But this one over here, it was like going against a brick wall, OK? It was not user-friendly, whatsoever.* [55-year-old NHW male]

It is important to note that subjective satisfaction concerns were limited and mentioned in only 2 of the 9 focus groups. Participants who successfully accessed the site's features and tasks enjoyed it. Participants across focus groups highlighted several positive aspects of the website:

> *[Diabetes MAP] reinforced some of the things I knew, but also gave me some new information, so I thought that was very good—I really enjoyed the educational features. I thought they were very helpful, [such as] what [medications do] to your body [and] how to take your medications.* [55-year-old African American/Black female]

> *I liked the skills [section], where it showed those 4 people and their tips on when to take the medicine.* [61-year-old African American/Black female]

> *I liked the My [Tailored] Tools part—I liked being able to read about you know, the consequences of not taking care of yourself when you have diabetes…they had some tips that I found…helpful. I love the text message notifications—that has increased my compliance.* [27-year-old NHW female]

Participant Recommendations

When participants voiced concerns about Diabetes MAP, they also gave suggestions for improving it. Common across all suggestions was a request for more simplicity and flexibility within the site. To improve the site's ease of learning and

efficiency of use, participants wanted more straightforward methods for accessing the site's components. They suggested strategically placing instructions, features, and Web content in easily recognizable ways. Some participants wanted specific features to be accessible on every page, or clearly designated on their own page. For example, participants liked the idea of a page dedicated to setting up text message reminders. On this page, previously entered medication information (ie, name, dosage) would appear and users could set up reminders for when to take specific medications and order prescription refills.

To improve the site's intuitive design, participants recommended the navigational videos be minimized or eliminated entirely.

> *[The navigational video] is a distraction for me because if I have to scroll down for whatever I have to do, it would be better for me if the video came up once you click it, and is then minimized.* [43-year-old Asian male]

Additionally, in reference to the site's memorability, participants recommended using a simple and recognizable URL that is easy to locate with an online search. They also wanted Diabetes MAP's purpose to be clear while using it (eg, spelling out the Diabetes MAP acronym and including images of diabetes medications throughout the website).

Participants wanted a simplified, streamlined Diabetes MAP user experience. In order to improve the site's error frequency and severity, participants recommended increasing compatibility across multiple browsers, including older versions of commonly used Web browsers. Additionally, they requested the ability to use Diabetes MAP across multiple digital devices without loading time delays. They also stressed the importance of clear and accessible resources for user support (eg, the ability to contact study staff directly if they had issues or questions, an accessible and searchable help resource on the website itself). Finally, participants wanted functional, useful, and easy to recall navigational videos to further facilitate learning how to use the site.

Those who reported some dissatisfaction, but generally endorsed the utility of Diabetes MAP, felt it might be more appropriate for certain types of patients with diabetes. For example, some users diagnosed with diabetes for a longer period of time felt the website might be particularly helpful for newly diagnosed patients. Other users suggested the website might be more useful for younger, more technology-proficient patients who prefer technology-delivered information as opposed to more traditional print materials.

Discussion

Principal Results

Usability testing is issue-focused and designed to assess the extent to which users can easily, efficiently, and effectively perform tasks with a technical system. We employed a mixed-methods approach to understand the challenges of using a Web-delivered medication adherence promotion intervention called Diabetes MAP. Participants with diabetes provided ratings and descriptions of their experiences using the website, as well as recommendations for improving it. On surveys, participants agreed Diabetes MAP was helpful and easy to use, but, in focus groups, they mentioned 24 unique user concerns related to each of the 6 factors determining website usability [22].

Our quantitative results are comparable to other usability studies employing the CSUQ, in that total ratings were above average [36,37]. When comparing survey items rated most to least favorably with focus group comments, there were instances when survey ratings and quotes were discordant and concordant. For example, participants rated the understandability and clarity of the site's information and the pleasantness of interacting with the site most favorably on surveys, yet, in focus groups, several participants expressed frustration with understanding how to complete tasks and navigate the website (ie, issues with the site's ease of learning and intuitive design). Alternatively, many positive statements about the value of the site's information and features (ie, subjective satisfaction) support these high ratings. Participants rated the ease and quickness of recovering from mistakes least favorably on surveys; focus group comments about error frequency and severity align with this low rating. In reconciling these inconsistencies and consistencies, it appears that while some participants had issues with understanding, navigating, and accessing Diabetes MAP, participants who successfully accessed the site, said it was enjoyable and helpful.

Recent usability studies of Web-delivered interventions for T2DM self- management yield results comparable to ours. In their evaluation of a Web-based dietary intervention, Ramadas et al found positive ratings of a website's usability based on survey items; however, this study did not use qualitative assessments [38]. Alternatively, Yu et al used focus groups to examine the usability of their self-management website [39] and identified several of the same usability concerns we did with Diabetes MAP. Namely, participants mentioned issues with the website layout and organization, navigation, data entry, and language [39]. Although these issues can be applied to Web-delivered interventions generally, usability testing also reveals issues specific to a certain website [39].

Our research highlights the value of using mixed methods for usability testing. Had we relied on only survey data, we would have incomplete information on Diabetes MAP's usability. Collecting qualitative data as part of usability testing reveals insights on unanticipated challenges and ideas for improving a site [21]. Additionally, involving members from the target audience is critical to understanding any unique needs of users for whom the site is intended [40]. The total time logged into Diabetes MAP during the 2-week period varied considerably across users. In a similar usability study, Heinrich et al had participants use a diabetes education site for 2 weeks, and participants visited the site an average of 3.6 ± 2.7 times and spent an average of 58.0 ± 56.1 total minutes on the site [41]. Compensation for time spent on the site was not reported. In our study, the more time spent using Diabetes MAP may reflect compensating participants per hour of use. Despite this, our qualitative results suggest some participants were discouraged from logging in more often because of the usability issues they encountered.

The US Department of Health and Human Services (HHS) has set forth peer-reviewed guidelines for improving the design and

usability of websites [42]. Taking into consideration these guidelines and the results of our usability study, we identified key principles for website creation to promote an optimal user experience. First, it is crucial to employ both simplicity and clarity in Web design. Diabetes MAP users were often confused about how to complete tasks and navigate the site both because they encountered user errors and the layout was unintuitive. HHS suggests standardizing tasks to be performed in a similar way so tasks can be reliably repeated [42]. When requesting users to enter information, a standard entry format should be used across tasks (eg, drop-down boxes). Additionally, to account for working memory limitations, content from 1 page that need be remembered on other pages should carry over to those pages [42]. Finally, using simplified and familiar terminology for a URL and website features will minimize user confusion and frustration.

Our second principle is to design websites with the goal of keeping users informed and aware of website processes. As a general observation from our focus groups, users became frustrated with unanticipated incidents (eg, long downloading times, automatically being logged out). In some cases, it may not be possible to reduce the size of a page to minimize the time it takes for a webpage to load [42]. However, a website can notify users of the time required to download an image and/or supply progress indicators (eg, an hourglass) to communicate a waiting period and its duration [42]. In either case, the user expects additional time instead of wondering how long to wait. HHS also recommends warning users if a page is going to "time out," so they can request extra time if needed [42]. Websites should also provide assistance to users who need additional help and ensure users are aware of this assistance. Resources should be easily accessible on the site, such as links providing more information about site content and a section for frequently asked questions [42].

Limitations

There are limitations to our study. Because we recruited our sample from a single academic medical center, our findings may not generalize to other patient populations. However, our sample characteristics map onto the academic medical center's patient population for whom Diabetes MAP was designed for. Additionally, although we were able to track time spent using Diabetes MAP, we were unable to track how participants used their Vanderbilt patient portal account because the two websites are not integrated. Considering the recent advancement of patient-provider communication in Web-delivered interventions, it will be valuable to track usage with this type of feature in future studies. Another limitation is the reliance on retrospective self-reports of users' experiences with Diabetes MAP. Furthermore, the range from when a participant used Diabetes MAP to when he/she participated in a focus group session was 14-88 days. Other usability testing methods, such as think-aloud protocols, cognitive walk-throughs, and remote user testing facilitate real-time data collection of user interactions with a system [30]. No single usability evaluation method can capture all usability problems [43]. While think-aloud studies combat the limitations of retrospective studies, they are limited by an unnatural situation in which users may feel uncomfortable talking to themselves and the possibility many statements will be filtered (ie, not reflective of users' actual experience) [44]. Although user feedback in our study was retrospective and the time between participants' website use and focus group feedback ranged widely, we presented Diabetes MAP on a large projector screen and re-oriented participants to its pages, functions, and features to solicit real-time feedback and prompt recall of users' experience. This method is similar to the pluralistic walkthrough approach that reveals users' uncertainty with a system's features and tasks better than other usability methods [31]. Additionally, the number of participants in each focus group allowed for the majority (over 80%) of concerns to be reported after only 2 focus group sessions. Moreover, the use of a mixed-methods approach provided a comprehensive evaluation of Diabetes MAP's usability. Finally, because usability feedback is system-dependent and the current study is based on a specific website, our findings may not generalize to other Web-delivered interventions.

Conclusions

Our findings highlight the importance of evaluating a website's usability prior to testing its efficacy. For Web-delivered interventions to be used as intended, researchers, Web designers, and developers must plan sufficient time to perform usability testing [45]. Ideally, they should begin with design thinking to allow for experimentation, creation and prototyping, and feedback and redesign prior to final product presentation [46]. An understanding of the target group's needs paired with iterative idea generation and development helps ensure the final website appeals to users and supports their using it [46]. Usability evaluation methods, including user-driven approaches, ensure users can optimally engage with and benefit from Web-delivered interventions [15,47]. Without ensuring a website's usability, critical design and functionality issues may go unnoticed and unaddressed, thereby preventing a site's benefits from being realized [15].

Acknowledgments

This research was funded with support from the Center for Diabetes Translational Research (NIDDK P30 DK092986) and CYO's Career Development Award (NIDDK K01 DK087894). CYO and LAN are both supported by NIDDK R01 DK100694.

Author LAN wrote and edited the manuscript. Author MCB co-designed the study, facilitated focus group discussions, collected, coded, and managed data, and conducted analyses. Author AL co-wrote and edited the manuscript. CYO co-designed the study, oversaw data collection and analyses, and co-wrote and edited the manuscript. The corresponding author and guarantor of the manuscript (CYO) takes full responsibility for the work as a whole, including the study design, access to data, and the decision to submit and publish the manuscript.

Conflicts of Interest

None declared.

References

1. Kirkman MS, Rowan-Martin MT, Levin R, Fonseca VA, Schmittdiel JA, Herman WH, et al. Determinants of adherence to diabetes medications: findings from a large pharmacy claims database. Diabetes Care 2015 Apr;38(4):604-609. [doi: 10.2337/dc14-2098] [Medline: 25573883]
2. Aikens JE, Piette JD. Longitudinal association between medication adherence and glycaemic control in Type 2 diabetes. Diabet Med 2013 Mar;30(3):338-344 [FREE Full text] [doi: 10.1111/dme.12046] [Medline: 23075262]
3. Jha AK, Aubert RE, Yao J, Teagarden JR, Epstein RS. Greater adherence to diabetes drugs is linked to less hospital use and could save nearly $5 billion annually. Health Aff (Millwood) 2012 Aug;31(8):1836-1846 [FREE Full text] [doi: 10.1377/hlthaff.2011.1198] [Medline: 22869663]
4. Ho PM, Rumsfeld JS, Masoudi FA, McClure DL, Plomondon ME, Steiner JF, et al. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med 2006 Sep 25;166(17):1836-1841. [doi: 10.1001/archinte.166.17.1836] [Medline: 17000939]
5. Currie CJ, Peyrot M, Morgan CL, Poole CD, Jenkins-Jones S, Rubin RR, et al. The impact of treatment noncompliance on mortality in people with type 2 diabetes. Diabetes Care 2012 Jun;35(6):1279-1284 [FREE Full text] [doi: 10.2337/dc11-1277] [Medline: 22511257]
6. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
7. Duncan I, Ahmed T, Li QE, Stetson B, Ruggiero L, Burton K, et al. Assessing the value of the diabetes educator. Diabetes Educ 2011;37(5):638-657. [doi: 10.1177/0145721711416256] [Medline: 21878591]
8. Perrin A, Duggan M. Pew Research Center. 2015. Americans' Internet Access URL: http://www.pewinternet.org/files/2015/06/2015-06-26_internet-usage-across-demographics-discover_FINAL.pdf [accessed 2015-06-26] [WebCite Cache ID 6babxjv1D]
9. Rosal MC, Heyden R, Mejilla R, Capelson R, Chalmers KA, Rizzo DM, et al. A virtual world versus face-to-face intervention format to promote diabetes self-management among african american women: a pilot randomized clinical trial. JMIR Res Protoc 2014 Oct;3(4):e54 [FREE Full text] [doi: 10.2196/resprot.3412] [Medline: 25344620]
10. Connelly J, Kirk A, Masthoff J, MacRury S. The use of technology to promote physical activity in type 2 diabetes management: a systematic review. Diabet Med 2013 Dec;30(12):1420-1432. [doi: 10.1111/dme.12289] [Medline: 23870009]
11. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an Internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
12. Lustria Mia Liza A, Noar SM, Cortese J, Van Stee Stephanie K, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
13. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009 Feb;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]
14. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
15. Yu CH, Bahniwal R, Laupacis A, Leung E, Orr MS, Straus SE. Systematic review and evaluation of web-accessible tools for management of diabetes and related cardiovascular risk factors by patients and healthcare providers. J Am Med Inform Assoc 2012;19(4):514-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000307] [Medline: 22215057]
16. Ramadas A, Quek KF, Chan C K Y, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
17. Brouwer W, Kroeze W, Crutzen R, de NJ, de Vries Nanne K, Brug J, et al. Which intervention characteristics are related to more exposure to Internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
18. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
19. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313 [FREE Full text] [doi: 10.1038/oby.2009.242] [Medline: 19696764]

20. O'Brien H, Toms E. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
21. Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. Curr Diab Rep 2014 Oct;14(10):534. [doi: 10.1007/s11892-014-0534-9] [Medline: 25173689]
22. U.S. Department of Health and Human Services (Digital Communications Division). Usability evaluation basics URL: http://www.usability.gov/what-and-why/usability-evaluation.html [accessed 2015-09-16] [WebCite Cache ID 6badPLndm]
23. Osborn CY, Mulvaney SA. Development and feasibility of a text messaging and interactive voice response intervention for low-income, diverse adults with type 2 diabetes mellitus. J Diabetes Sci Technol 2013 May;7(3):612-622 [FREE Full text] [Medline: 23759393]
24. Mayberry LS, Osborn CY. Empirical validation of the information-motivation-behavioral skills model of diabetes medication adherence: a framework for intervention. Diabetes Care 2014;37(5):1246-1253 [FREE Full text] [doi: 10.2337/dc13-1828] [Medline: 24598245]
25. Osborn CY, Egede LE. Validation of an Information-Motivation-Behavioral Skills model of diabetes self-care (IMB-DSC). Patient Educ Couns 2010 Apr;79(1):49-54 [FREE Full text] [doi: 10.1016/j.pec.2009.07.016] [Medline: 19699601]
26. Fisher JD, Amico KR, Fisher WA, Cornman DH, Shuper PA, Trayling C, et al. Computer-based intervention in HIV clinical care setting improves antiretroviral adherence: the LifeWindows Project. AIDS Behav 2011 Nov;15(8):1635-1646. [doi: 10.1007/s10461-011-9926-x] [Medline: 21452051]
27. Nielsen J. Nielsen Norman Group. 2000. Why You Only Need to Test with 5 Users URL: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2015-09-15] [WebCite Cache ID 6badu0T87]
28. Nielsen J. Nielsen Norman Group. 2006. Quantitative Studies: How Many Users to Test? URL: http://www.nngroup.com/articles/quantitative-studies-how-many-users/ [accessed 2015-09-28] [WebCite Cache ID 6bt6qz15W]
29. Harris P. Research Electronic Data Capture (REDCap)-planning, collecting and managing data for clinical and translational research. BMC Bioinformatics (Suppl 12) 2012;13:A15.
30. Nielsen J. Nielsen Norman Group. 1995. Summary of Usability Inspection Methods URL: http://www.nngroup.com/articles/summary-of-usability-inspection-methods/ [accessed 2015-09-16] [WebCite Cache ID 6bagIeU09]
31. Riihiaho S. The pluralistic usability walk-through method. Ergon Des 2002;10(3):23-30.
32. Lewis J. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Int 1995;7(1):57-78.
33. Strauss A, Corbin J. Basics of qualitative research: Techniques and procedures for developing grounded theory, 2nd ed. Thousand Oaks, CA: Sage Publications; 1998.
34. Saldaña J. The Coding Manual for Qualitative Researchers, 2nd ed. Thousand Oaks: California Sage Publications Inc; 2012:978-971.
35. Vanderbilt Institute for Medicine and Public Health. Vanderbilt Health. Vanderbilt University Hospitals and Clinics: Community Health Needs Assessment URL: http://www.mc.vanderbilt.edu/documents/main/files/VU%20CHNA%20Report%202013.pdf [accessed 2015-06-26] [WebCite Cache ID 6fqJrr1kg]
36. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
37. Wismer A, Madathil K, Koikkara R, Juang K, Greenstein J. Evaluating the usability of CAPTCHAs on a mobile device with voice and touch input. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting.: SAGE Publications; 2012 Presented at: Human Factors and Ergonomics Society Annual Meeting; 2012; Boston.
38. Ramadas A, Chan Carina Ka Yee, Oldenburg B, Hussien Z, Quek KF. A web-based dietary intervention for people with type 2 diabetes: development, implementation, and evaluation. Int J Behav Med 2015 Jun;22(3):365-373 [FREE Full text] [doi: 10.1007/s12529-014-9445-z] [Medline: 25274015]
39. Yu CH, Parsons JA, Hall S, Newton D, Jovicic A, Lottridge D, et al. User-centered design of a web-based self-management site for individuals with type 2 diabetes - providing a sense of control and community. BMC Med Inform Decis Mak 2014;14:60 [FREE Full text] [doi: 10.1186/1472-6947-14-60] [Medline: 25056379]
40. Kinzie MB, Cohn WF, Julian MF, Knaus WA. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. J Am Med Inform Assoc 2002;9(4):320-330 [FREE Full text] [Medline: 12087113]
41. Heinrich E, de NJ, Schaper NC, Schoonus-Spit Maartje H G, Janssen Monique A J, de Vries Nanne K. Evaluation of the web-based Diabetes Interactive Education Programme (DIEP) for patients with type 2 diabetes. Patient Educ Couns 2012 Feb;86(2):172-178 [FREE Full text] [doi: 10.1016/j.pec.2011.04.032] [Medline: 21616626]
42. U.S. Department of Health and Human Services. The Research-Based Web Design & Usability Guidelines, Enlarged/Expanded edition. Washington: U.S. Government Printing Office; 2006. URL: http://www.usability.gov/sites/default/files/documents/guidelines_book.pdf [accessed 2015-06-25] [WebCite Cache ID 6bt6tSPsC]
43. Fernandez A, Insfran E, Abrahão S. Usability evaluation methods for the web: a systematic mapping study. Inform Software Tech 2011;53(8):789-817. [doi: 10.1016/j.infsof.2011.02.007]
44. Nielsen J. Nielsen Norman Group. 2012. Thinking Aloud: The #1 Usability Tool URL: http://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ [accessed 2015-06-25] [WebCite Cache ID 6fqJDWCnu]

45. Henderson VA, Barr KL, An LC, Guajardo C, Newhouse W, Mase R, et al. Community-based participatory research and user-centered design in a diabetes medication information and decision tool. Prog Community Health Partnersh 2013;7(2):171-184 [FREE Full text] [doi: 10.1353/cpr.2013.0024] [Medline: 23793248]
46. Razzouk R, Shute V. What is design thinking and why is it important? Rev Educ Res 2012 Aug 09;82(3):330-348. [doi: 10.3102/0034654312457429]
47. Belanche D, Casaló LV, Guinalíu M. Website usability, consumer satisfaction and the intention to use a website: the moderating effect of perceived risk. Journal of Retailing & Consumer Services 2012 Jan;19(1):124-132. [doi: 10.1016/j.jretconser.2011.11.001]

Abbreviations

A1c: hemoglobin A1c
CSUQ: Computer System Usability Questionnaire
Diabetes MAP: Diabetes Medication Adherence Promotion
IMB: Information-Motivation-Behavioral skills
HbA1c: glycated hemoglobin A1c
HHS: US Department of Health and Human Services
RA: research assistant
REDCap: Research Data Capture
SD: standard deviation
T2DM: type 2 diabetes mellitus
NHW: non-Hispanic White

Edited by P Morita; submitted 28.09.15; peer-reviewed by L Tieu, J MacSharry, K Barnard, S Goyal, K Linden; comments to author 22.10.15; revised version received 30.11.15; accepted 27.01.16; published 12.05.16.

Please cite as:
Nelson LA, Bethune MC, Lagotte AE, Osborn CY
The Usability of Diabetes MAP: A Web-delivered Intervention for Improving Medication Adherence
JMIR Human Factors 2016;3(1):e13
URL: http://humanfactors.jmir.org/2016/1/e13/
doi:10.2196/humanfactors.5177
PMID:27174496

Original Paper

Effects of Information Visualization on Older Adults' Decision-Making Performance in a Medicare Plan Selection Task: A Comparative Usability Study

Margaux M Price[1], PhD; Jessica J Crumley-Branyon[2], MS; William R Leidheiser[2], MS; Richard Pak[2], PhD

[1]John Deere, Urbandale, IA, United States

[2]Clemson University, Department of Psychology, Clemson, SC, United States

Corresponding Author:
Richard Pak, PhD
Clemson University
Department of Psychology
418 Brackett Hall
Clemson, SC, 29634
United States
Phone: 1 864 656 1584
Fax: 1 864 656 0358
Email: richpak@clemson.edu

Abstract

Background: Technology gains have improved tools for evaluating complex tasks by providing environmental supports (ES) that increase ease of use and improve performance outcomes through the use of information visualizations (info-vis). Complex info-vis emphasize the need to understand individual differences in abilities of target users, the key cognitive abilities needed to execute a decision task, and the graphical elements that can serve as the most effective ES. Older adults may be one such target user group that would benefit from increased ES to mitigate specific declines in cognitive abilities. For example, choosing a prescription drug plan is a necessary and complex task that can impact quality of life if the wrong choice is made. The decision to enroll in one plan over another can involve comparing over 15 plans across many categories. Within this context, the large amount of complex information and reduced working memory capacity puts older adults' decision making at a disadvantage. An intentionally designed ES, such as an info-vis that reduces working memory demand, may assist older adults in making the most effective decision among many options.

Objective: The objective of this study is to examine whether the use of an info-vis can lower working memory demands and positively affect complex decision-making performance of older adults in the context of choosing a Medicare prescription drug plan.

Methods: Participants performed a computerized decision-making task in the context of finding the best health care plan. Data included quantitative decision-making performance indicators and surveys examining previous history with purchasing insurance. Participants used a colored info-vis ES or a table (no ES) to perform the decision task. Task difficulty was manipulated by increasing the number of selection criteria used to make an accurate decision. A repeated measures analysis was performed to examine differences between the two table designs.

Results: Twenty-three older adults between the ages of 66 and 80 completed the study. There was a main effect for accuracy such that older adults made more accurate decisions in the color info-vis condition than the table condition. In the low difficulty condition, participants were more successful at choosing the correct answer when the question was about the gap coverage attribute in the info-vis condition. Participants also made significantly faster decisions in the info-vis condition than in the table condition.

Conclusions: Reducing the working memory demand of the task through the use of an ES can improve decision accuracy, especially when selection criteria is only focused on a single attribute of the insurance plan.

(JMIR Hum Factors 2016;3(1):e16) doi:10.2196/humanfactors.5106

KEYWORDS

Information visualization; aging; health-related websites; working memory

Introduction

Older Adults' Difficulties With the Medicare Website

A usability evaluation of the Medicare website revealed that older adults were unable to successfully choose a prescription drug plan for a given medication regimen [1]. Example problems highlighted in the evaluation included general difficulties navigating the site, frustration, and the inability to locate desired information. Compared to younger age groups, older adults have less success obtaining Internet health information [2]. Insurance and medical jargon (eg, "gap coverage", "drug sharing", etc) may have further exacerbated the difficulties. Even without time constraints, difficulties in identifying the best plan in a demanding environment can lead users to select a plan that does not provide adequate medical coverage or is more expensive than other available options [3].

The number of prescription drug plans presented to users can cause severe problems, especially when users attempt to simultaneously compare choices across different criteria. Simply increasing the number of available Medicare drug plans from 3 to 9 is associated with poorer decision outcomes that can negatively affect quality of life, quality of care, and overall health [4]. Poorer decision outcomes result from the increased working memory demands associated with comparing a larger number of plans. Trying to make optimal decisions in the face of uncertainty with a large amount of inputs is a working memory demanding task (see Multimedia Appendix 1). In particular, comparing plans across different criteria and calculating costs (steps 5-9) illustrates the working memory-intensive nature of selecting an appropriate prescription drug plan. The working memory-intensive nature of the task combined with older adults' reduced capacity for working memory [5] result in a reduced ability to discern between plan costs as the number of plans increase compared to their higher cognitive functioning counterparts [6]. In sum, choosing an optimal health plan is a task that places large demands on working memory and attention and can result in negative consequences if decision-making performance suffers. A decision aid designed to redirect task demands from working memory to an external memory aid may facilitate optimal decision-making in older adults.

The Importance of Working Memory in Decision-Making

Working memory capacity refers to the amount of information one can temporarily store and manipulate at any given time [7]. If a task's working memory demand exceeds one's working memory capacity, then task performance declines. This capacity limit is central to one's ability to process information and make a decision. The information processing model of decision-making [8] is a useful tool for understanding working memory demands at each step of the decision-making process. According to the model, information must first be perceived then selectively attended to by the decision maker. Next, the decision maker generates hypotheses about specific outcomes and selects a decision or response. Finally, the decision maker implements the response and compares the outcome to the initial set of hypotheses. Each step of this model as it relates to choosing a prescription drug plan is discussed in further detail below.

Attentional limitations force the user to filter cues relevant to the decision goal from the irrelevant cues by selectively attending to only some of the information present. Choosing a prescription drug plan on the basis of cost first requires that the decider perceive the appropriate cues (eg, monthly premiums, coverage in the gap), while ignoring irrelevant cues (eg, Medicare ID numbers or contact information). Cues are selected based on their diagnosticity (amount of information the cue provides), reliability (trustworthiness of information), and salience (physical properties such as volume, color, and shape). After cues are selected for further processing, they are compared to other information to form a meaningful interpretation of the state of the system.

After selectively attending to appropriate cues, the information is manipulated in working memory where hypotheses or potential outcomes are generated (eg, plans with a low monthly premium and low deductibles have less coverage). Choosing a prescription drug plan requires several hypotheses for each plan; one for cost and the effect on personal budget, one for satisfaction, etc. Here, working memory limitations prevent a truly exhaustive comparison. The next step involves integrating the outcomes and action selection. At this stage, the decision maker tries to determine which option will produce an outcome that best meets the goal. If a plan is selected for its low monthly premium, but also has a low satisfaction rating, the decision maker has to consider the potential implications of both attributes together. This step is highly error-prone because working memory capacity limits the number of comparisons that can be made simultaneously. When an action is selected and carried out (a decision is made), the outcome is monitored and evaluated against new cues or information, and new hypotheses about the state of the system are formed. Here again, working memory capacity limits the amount of new information that can be selected and then compared to the current state of the system, potentially harming the ability to select the best plan.

Effects of Aging on Decision Making

Older adults' reduced working memory capacity [5] limits the number of integration and comparison tasks that can be made at a given time and thus will affect their ability to make optimal decisions [9]. Age-related limitations may force older adults to rely more heavily on heuristic-based decision making (ie, "rules of thumb" or cognitive shortcuts used to make decisions quickly with little effort) [10], which may not always lead to an optimal decision. Although older adults are sometimes successful in adapting their strategies to meet task demands, they tend to perform worse on the task of integrating information (comparing more than two pieces of information) and extracting information (finding one piece of information) [11]. For example, comparing information that is presented in different units (eg, monetary units and satisfaction ratings) could make a task more difficult for older adults [12,13]. Indeed, when choosing a prescription drug plan, one must compare multiple cost values and multiple satisfaction ratings across many different plans.

Older adults also tend to commit more errors and have more difficulty comprehending information than younger adults when the task requires integrating information [12] among many choices [13]. One way to reduce errors, besides reducing the number of possible choices, is to include specific visual aids that guide attention toward more relevant choices and help eliminate the need to hold information about less relevant choices in working memory.

Although working memory limits the amount of information used to make optimal decisions, information that shares similar perceptual or semantic features may be grouped together into object-like “chunks” or visual clusters that enable pattern recognition [14-16], effectively overcoming some working memory limitations. Information may be chunked together based on color, shape, meaning, spatial proximity, or other properties (eg, Gestalt principles) [17] preattentively or automatically (without the need to selectively attend to each cue individually). This perceptual integration process may help facilitate processing of more information with less effort.

Aids that reduce working memory demands are called environmental supports (ES) [18]. ESs often utilize perceptual integration principles to improve task performance for older adults by reducing task demands or supporting the use of existing resources [19,20]. Several studies with younger adults have shown that providing an ES reduces working memory demand by facilitating visual search and automatic perceptual processing of information from graphs [15] by visually integrating related information into meaningful chunks using color [21]. Ratwani et al [15] theorized that when information within a graph is organized into visual “clusters”, less effort is needed to group similar information together, which reduces the working memory demand of the task; the user can focus attention on the differences between the groups, rather than first actively integrating information into clusters.

Reducing the need for effortful comparison of information will allow the user to allocate more resources to later steps in the decision-making process, which could result in more thorough outcome predictions (eg, how a plan might affect finances) and appropriate action selections (eg, choosing one plan over another) [15]. Older adults may benefit from a decision aid designed to shift task demands from working memory to an external memory aid [22], where it can be perceived by the relatively age-insensitive preattentive visual perceptual system [23]. For example, an ordered brightness scale allows people to make comparisons between choices without having to process a number and assign it meaning before serially moving on to the next choice [24]. Instead, meaning is automatically processed using perceptual features (eg, darker green may represent a higher number than a lighter green, the scale is based on the color density). Additionally, it is much faster to search for a color singleton than to find a number target [25]. This suggests a promising avenue of providing an ES-based decision-making aid: shifting the working memory burden to the perceptual processing system by eliminating the need to comprehend and compare each option semantically, and instead comparing the information perceptually.

Objective

The objective of the current study was to extend Lohse's [21] and Ratwani et al's [15] findings to the design of an information visualization (info-vis) aid in order to examine whether older adult decision-making performance can be enhanced by the use of graphical decision aids designed to reduce working memory demands. Reducing working memory demands was expected to lessen reliance on less accurate heuristic strategies and improve decision quality. Decision quality was measured by how well the choice met the criterion in the question. The assumption was that when the decision-making task is reduced from cognitively complex to relatively easy, decision makers would not need to rely on heuristics and would consider all relevant information. Specifically, we predict an interaction between decision aid and task difficulty such that the more difficult task decisions will be benefitted most by the info-vis aid. Because the info-vis aid was designed to reduce working memory demands of the task, the performance gains will be greater for difficult tasks that require more working memory resources.

Methods

Participants

There were 23 older participants ages 65-80 that were recruited through an existing database of volunteers in the surrounding community. Older adults received US $14 in compensation for participating. Color-blindness and the inability to read a computer screen were the only exclusion criteria.

Design

The study was a 2 (decision aid: table, color info-vis) x 2 (task difficulty: low, high) repeated measures design, with decision aid as the between-subjects variable and task difficulty as the within-subjects variable. Therefore, each participant was randomly assigned to one of the decision aid conditions and completed trials at both levels of task difficulty. Participants made decisions on a total of 20 trials. The trials were organized around 4 blocks of 5 questions per block. A randomized blocked design was utilized for questions of varying task difficulty. The questions within each block were also randomly presented. Dependent measures included decision accuracy (sum score of number correct), decision quality (sum score of scaled decision ratings for the high difficulty questions), and decision task time (in seconds).

Materials

Demographic information, health information, insurance experience, technology experience, a working memory measurement, and an exit survey were collected from each subject. A blocked design allowed us to administer the subjective workload measure (NASA-TLX) at the end of each block for each level of task difficulty and working memory demand. Participants used PC-compatible computers and wore headphones during the experiment. The experiment was programmed using E-prime (version 1.1).

Task

Decision Task

The decision task utilized a computerized decision-making paradigm presented in the context of choosing the best health care plan based on given criteria. All participants were assigned to one of the two decision aid conditions and performed tasks at both levels of difficulty. A standardized format was used so that the question, plan data, and choice set always appeared in the same location for each trial. The question was located at the top of the screen, with the decision aid below it.

Decision Aids

The table condition was a replica of the table found on the Medicare website (as shown in 2010). The table included a row for each of the fifteen prescription drug plans and columns for four of the plan's attributes (Figure 1 shows this). In 2014 and 2015, an average of eighteen Medicare Advantage plans were available to enrollees [26]; therefore fifteen plans are representative of a typical Medicare plan selection task.

The info-vis condition was created by adding graphics instead of (or in addition to) text to represent specific attributes. Visualization tools are able to help users interpret large sets of information quickly and efficiently [27]. A single info-vis (Figure 2 shows this) was created utilizing well-accepted display design principles (eg, proximity compatibility principle, color gradients, pictorial representations, and redundancy) [28,29]. The info-vis used in this study was specifically designed to alleviate the working memory intensive parts of the task by converting them into easier perceptual tasks using a color manipulation. Color info-vis uses multi-colored scales (heat map color scale) to replace the categorical gap coverage text. The same multi-colored scale was used in the stars that replace the number scales for satisfaction ratings. The colors highlight the relevant information within each attribute and create fewer mental comparisons for the user. Multi-colored scales can facilitate identification tasks—where one has to select a target value represented by a color (eg, identify the plans that have gap coverage level of all generics—represented by the color green), and in cases where a particular absolute value (ie, all generics) is more important than a relative value (ie, the plan with the lowest amount of coverage) [24]. In the current study, the multi-colored scale was used to represent the five specific categories of both gap coverage and satisfaction ratings and these categories were absolute, not relative to one another (eg, "all generics" was always the highest level of gap coverage, but "some" or "many" generics are not proportionate to each other).

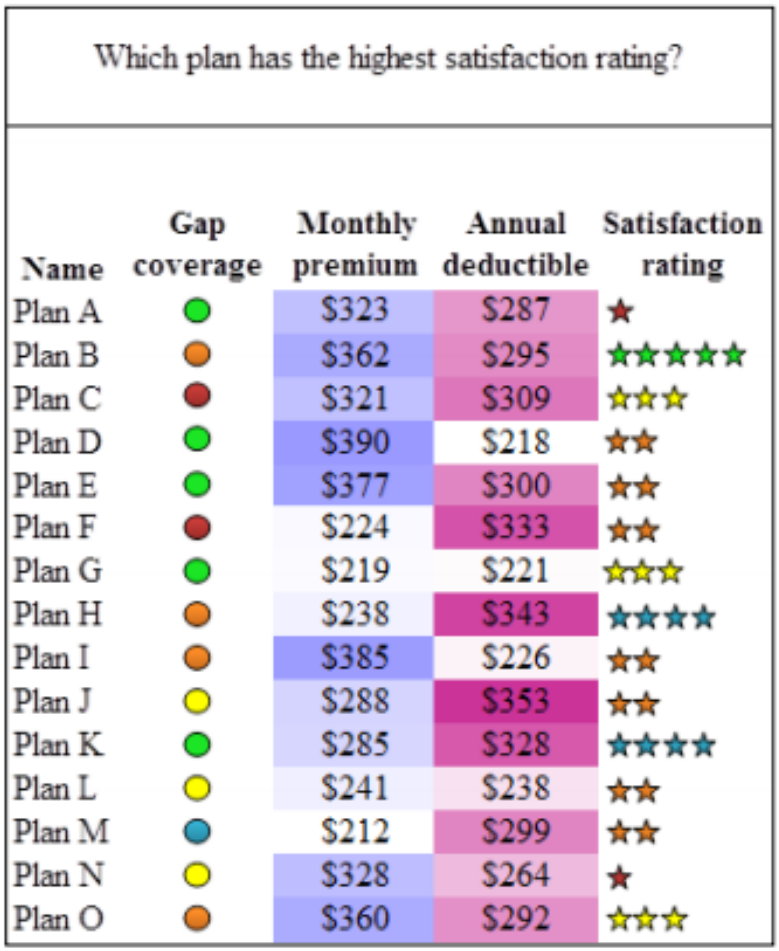
Brightness ordered scales (same color is used, but lightest color gradient is the lowest value and the darkest color is the highest value) were added to dollar amounts in both the monthly premium and annual deductible columns. Brightness ordered scales have been shown to be superior for comparisons of relative value [24] where all values are compared to one another (eg, which plan has the lowest or highest monthly premium). These color manipulations were added to facilitate more perceptual comparisons rather than effortful cognitive comparisons, thus reducing working memory demand. Each attribute in the display layout was grouped close together in perceptual space to make comparisons easier (ie, proximity compatibility principle) [29]. Pilot testing on a younger adult sample found that color info-vis does minimize working memory demand. Participants made more accurate decisions with a color info-vis than size or no info-vis. Size info-vis used pie and bar graphs to indicate differences between drug plans for each criterion. The pilot testing followed a similar procedure except that an auditory n-back task was used to constrain younger adults' working memory capacity. The secondary task simulated the limitations on older adults' cognitive abilities.

Figure 1. Example layout of a low difficulty decision task in the table condition. Fifteen plan options are shown with four plan attributes (gap coverage, monthly premium, annual deductible, and satisfaction rating).

Which plan has the highest satisfaction rating?

Name	Gap coverage	Monthly Premium	Annual Deductible	Satisfaction Rating
Plan A	All generics	$323	$287	1.0 out of 5 stars
Plan B	Some generics	$362	$295	5.0 out of 5 stars
Plan C	No gap coverage	$321	$309	3.0 out of 5 stars
Plan D	All generics	$390	$218	2.0 out of 5 stars
Plan E	All generics	$377	$300	2.0 out of 5 stars
Plan F	No gap coverage	$224	$333	2.0 out of 5 stars
Plan G	All generics	$219	$221	3.0 out of 5 stars
Plan H	Some generics	$238	$343	1.0 out of 5 stars
Plan I	Some generics	$385	$226	2.0 out of 5 stars
Plan J	Many generics	$288	$353	2.0 out of 5 stars
Plan K	All generics	$285	$328	4.0 out of 5 stars
Plan L	Many generics	$241	$238	2.0 out of 5 stars
Plan M	Most generics	$212	$299	2.0 out of 5 stars
Plan N	Many generics	$328	$264	1.0 out of 5 stars
Plan O	Some generics	$360	$292	3.0 out of 5 stars

Figure 2. Color information visualization (color info-vis).



Task Difficulty

Task difficulty was manipulated by varying the number of plan attributes that must be considered in order to accurately complete the task. In the low difficulty condition, participants selected a plan based on one attribute (eg, Which plan has the lowest monthly premium?). The high difficulty condition required the participant to select a plan by integrating and comparing three attributes of each plan (eg, Which plan has the lowest monthly premium, highest gap coverage, and highest satisfaction rating?).

For both conditions, the data were structured so that only one plan best met all of the criteria in the question. This manipulation required participants to make a compensatory decision (choosing the best plan by evaluating alternatives along with the required selection criteria) and use an analytical decision strategy in order to select the best answer [30], thus, using heuristics would not lead to the optimal answer choice. Participants in the low difficulty condition only had to compare the values for a single attribute. In the current info-vis table, each attribute is integrated with graphics that makes identifying the optimal choice for each attribute less cognitively demanding. he low difficulty condition is practically useful because single-attribute decision making is a common heuristic in naturalistic decision-making [30] and establishes a baseline of performance on which other conditions can be compared against. Analysis of a single-attribute decision will answer whether the graphical representation (info-vis) can also affect the efficiency and accuracy of identifying the best option based on a single attribute. In the high difficulty condition, participants needed to rank the values for each attribute and add them together to identify the best plan. The high difficulty condition approximates rational decision-making techniques that consider larger amounts of information and require greater cognitive demands.

Procedure

Experimental sessions were administered in groups of 1 to 4 participants; however, each participant worked independently. After providing informed consent, participants completed a paper and pencil working memory ability test, the Reverse Digit Span [31], before moving on to the computerized portion of the task.

The terms used in the decision task (eg, gap coverage) were defined by the experimenter and also presented visually on the screen. Participants first completed a series of practice questions that introduced low and high difficulty problems of the decision-making task. Participants chose an answer by pressing the letter on the keyboard that corresponded with the selected plan (eg, participants pressed the “A” key to select Plan A). At the end of the practice task, a screen prompted users to fill out the NASA-TLX survey. Participants then began the recorded trials. Each recorded block involved the same procedure as the practice block. At the conclusion of the task, participants completed the demographics and health survey, a technology

experience survey, an insurance purchasing experience questionnaire, and an exit survey.

Results

Participants

There were 23 older adults (12 female) between the ages of 66 and 80 (mean, M, 72.4, SD 3.73) that participated in this study. Many of the participants indicated they had previous experience purchasing insurance (ie, 19 out of 23 participants, 83%, bought Medicare plans, 14 out of 23, 61%, bought prescription drug insurance, and 20 out of 23, 87%, bought health insurance). No significant differences (P>.05) were found between decision aid groups on computer experience, health, insurance purchasing experience, working memory, or age. Therefore, all subjects were included in the following analyses.

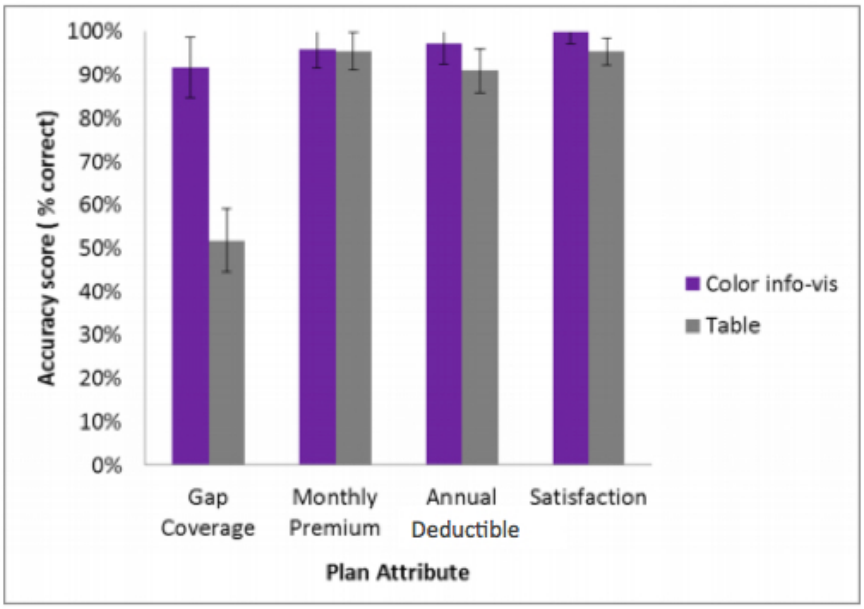
Decision Accuracy

A 2 (decision aid) x 2 (difficulty) repeated measures analysis of variance (ANOVA) revealed a significant main effect of task difficulty on decision accuracy ($F_{1, 21}$=39.88, P<.001, η^2=.65). Participants performed the decision task more accurately in the low difficulty condition (M 8.87, SD 1.39) than in the high difficulty condition (M 6.30, SD 2.05). There was a significant main effect of decision aid ($F_{1, 21}$=3.81, P=.03, η_p^2=.15), which confirms the hypothesis that older adults would perform significantly better in the color info-vis condition (M 8.13, SD 1.21) than the table condition (M 7.00, SD 1.55).

The interaction between task difficulty and decision aid was not significant ($F_{1, 21}$=.829, P=.19, η_p^2=.04).

For the low difficulty decision tasks, participants were asked to find a plan that best meets the single criterion (one attribute, eg, satisfaction rating). Therefore, we can analyze performance for each attribute (gap coverage, monthly premium, annual deductible, and satisfaction rating) individually to examine why participants were more accurate in the info-vis condition than in the table condition. Because the high difficulty condition involves locating a plan that meets several criteria, we can only assess the source of the main effect of decision aid in the low difficulty condition. The low difficulty condition data were analyzed using a 2 (decision aid) x 4 (plan attribute) mixed measures ANOVA. Main effects of attribute type ($F_{1.72, 36.11}$=15.61, P<.001, η_p^2=.43) and decision aid ($F_{1, 21}$=7.10, P=.02, η_p^2=.25) were qualified by a significant interaction between plan attribute and decision aid ($F_{1.72, 36.11}$=8.81, P=.001, η_p^2=.30). Figure 3 shows this. Participants were better able to accurately answer questions about the gap coverage attribute in the color info-vis condition (M 91.7%, SD 20.77%) than in the table condition (M 51.73%, SD 27.51%). This difference is the source of the main effect of decision aid on accuracy.

Figure 3. Percent accuracy on low difficulty tasks by plan attribute and decision aid. Error bars represent standard error of the mean.



Decision Task Time

A 2 (decision aid) x 2 (difficulty) repeated measures ANOVA was run to assess decision task time (in seconds, s) and revealed a significant main effect of difficulty ($F_{1, 21}$=155.73, P<001, η_p^2=.88), such that participants were faster in the low difficulty condition (M 20.07 s, SD 7.78 s) than in the high difficulty condition (M 70.69 s, SD 20.92 s). Figure 4 shows this. There was no significant main effect of decision aid ($F_{1, 21}$=1.07, P=.31, η_p^2=.05) on task time, nor was there an interaction between decision aid and difficulty ($F_{1, 21}$=.081, P=.78, η_p^2=.01).

A 2 (decision aid) x 4 (plan attribute) repeated measures ANOVA on decision time (in s) was run to look for evidence of a speed-accuracy trade-off that might explain the effect of decision aid on accuracy with gap coverage questions. We can only assess the source of the main effect in the low difficulty condition because the high difficulty condition involves locating a plan that meets several criteria. The analysis revealed a

significant main effect of decision aid ($F_{1, 21}$=4.5, P=.046, η_p^2=.18) and a significant main effect of plan attribute ($F_{6.8, 37.68}$=6.82, P=.004, η_p^2=.25), but not a significant interaction between decision aid and attribute (P=.08). Participants spent more time answering the gap coverage questions than the other attributes and more time answering questions about this attribute in the table condition than in the color info-vis condition (Figure 5 shows this).

Participants answered the decision task significantly faster in the color info-vis condition (M 16.93 s, SD 5.95 s) than in the table condition (M 23.5 s, SD 8.35 s). Questions about the satisfaction rating attribute (M 13.69 s, SD 8.81 s) took significantly less time than the annual deductible (M 19.64 s, SD 5.22 s), gap coverage (M 25.41 s, SD 17.66 s), and monthly premium (M 19.56 s, SD 6.86 s). This indicates that there was not a speed-accuracy trade-off that would explain significantly lower accuracy for gap coverage questions in the table condition versus the color info-vis condition.

Figure 4. Decision task time by decision aid for low and high difficulty tasks. Error bars represent standard error of the mean.

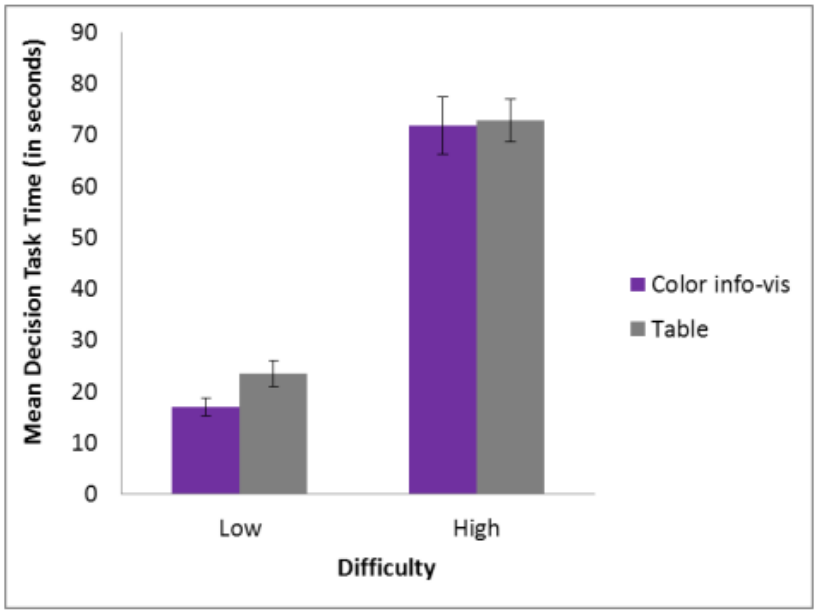
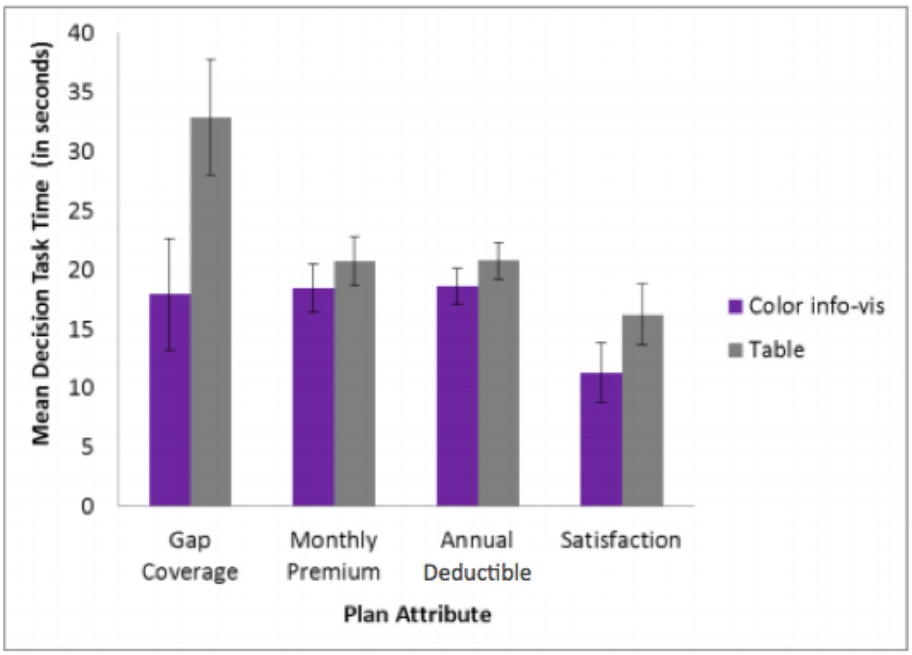


Figure 5. Mean decision time (in seconds) by plan attribute and decision aid for the low difficulty condition. Error bars represent standard error of the mean.



Decision Quality

For each high difficulty question, the plan data were created so that only one option met all of the criteria presented in the question during each trial. The other plan options met only 0, 1, or 2 of the 3 possible criterion. Choosing the correct plan assumes that each criterion was used in the assessment. Thus, a maximum score of 3 is possible for each question and

represents the best answer. A minimum score of 0 indicates that the plan chosen met none of the criteria in the question. These points were added together to compute a total decision quality score for the high difficulty questions. For the computed score, the maximum score was 30 points (3 x 10 questions) and the minimum score was 0 points.

An independent samples *t* test was conducted between decision aid conditions on decision quality score and revealed that quality did not differ significantly by decision aid (t=.7, P=.49). A one-tailed significance test did not change the effect of the decision aid variable on decision quality.

Subjective Workload

Subjective workload ratings were assessed by conducting a 2 (decision aid) x 2 (difficulty) repeated measures ANOVA. A significant main effect of difficulty ($F_{1,\ 21}$=74.2, P<.001, η_p^2=.78) was revealed in the expected direction. This was a manipulation check for difficulty and indicates a successful manipulation because participants rated the high difficulty tasks as significantly more difficult (M 58.63, SE 3.57) than the low difficulty tasks (M 35.35, SE 2.99). There was no main effect of decision aid ($F_{1,\ 21}$=1.5, P=.23, η_p^2=.07), nor an interaction effect of decision aid and difficulty ($F_{1,21}$=.06, P=.82, η_p^2=.003).

Discussion

Principal Findings

This study examined whether color info-vis can be used as a decision support for older adults making complex decisions. Previous research has shown that older adults exhibit difficulty in choosing a prescription drug plan on the Medicare website, possibly because of a combination of usability issues and normative changes in cognitive abilities such as reduced working memory capacity [1]. It was hypothesized that older adults would perform better (higher accuracy and quality) in the color info-vis condition than in the table condition for both high and low difficulty tasks. Our results show that accuracy was significantly higher in the color info-vis condition (shifting processing burden from cognitive resources to perceptual resources) than in the table condition, indicating that older adults did not use heuristics, but instead an analytical decision-making strategy.

If older adults did not choose the best plan option, they were able to select a plan that was "good enough" in quality regardless of the decision aid. This finding is consistent with the finding that older adults are more likely to use heuristic strategies at a lower level of working memory demand than younger adults and that they can use heuristics successfully [10].

It was hypothesized that performance in the difficult task condition would benefit most from the info-vis display. However, the interaction between task difficulty and decision aid was not significant. The lack of an effect of condition on accuracy in the high difficulty tasks indicates that relying on perceptual capacities cannot fully accommodate age-related declines in cognitive capacities (color info-vis condition). Although the color info-vis may have been successful in reducing the working memory demand for comparing plans on a single attribute (low difficulty task), the info-vis did little to support integration of more than one attribute (ie, the three attributes required in the high difficulty tasks). That is, the info-vis display used color to reduce working memory demands when making decisions within each attribute, but the table did not facilitate information integration or show relationships among different attributes. This could also account for the finding that the type of aid did not influence perceived workload. Future research should evaluate ways to support more complex decision-making tasks where multiple attributes must be integrated and compared through info-vis (eg, configural displays).

In the graph reading literature, a low difficulty condition is generally termed an extraction task because the user is asked to find a specific bit of information (eg, what is Plan B's monthly premium amount), rather than perform a comparison of one attribute among many options (eg, which plan has the lowest monthly premium) as in this study. This may be why there was an effect in the low difficulty condition that is not consistently found in other studies within the graph reading literature [15].

In the low difficulty condition, older adults were much more successful at choosing the correct answer when the question was about the gap coverage attribute. This finding is interesting for a number of reasons. First, although the performance boost in the gap coverage attribute is selective, it could be due to ceiling effects in the accuracy data, and floor effects in task time data. The gap coverage attribute had the most room for improvement among the other attributes in both accuracy and task time. In the table condition, accuracy for the monthly premium, annual deductible, and satisfaction ratings attributes all approached near optimal levels of accuracy, while gap coverage yielded less accurate responses. There was similar room for improvement in the task time data such that task times in the table condition for monthly premium, annual deductible, and satisfaction rating were around 20 s compared to approximately 33 s for the gap coverage decision. Second, the user had to remember what each of the colors meant or had to refer to the legend, which on the surface appears to increase working memory demand. However, in the table condition, gap coverage had to be evaluated based on textual values (eg, all generics vs some generics). This requires reading and comprehension of the text, rather than a less working memory demanding visual search for a target color [25]. Third, previous literature has suggested that numeracy (ability to process numerical information) and processing speed (or how fast one can process information and perform tasks without focused attention) are responsible for performance differences with a large dataset (24 plan options) [13]. Using color comparisons rather than numerical comparisons may be a good option for those who do not have high numeracy abilities, working memory abilities, and those with slower processing speed.

Whether or not the use of color is in fact allowing the user to make faster, less demanding comparisons might be a question that can be answered using eye-tracking data. For example, recording fixation durations and plotting saccadic amplitude could help answer the question of whether color is facilitating

a less cognitively demanding search [32]. Long fixation durations might indicate focal vision, which is indicative of selective attention, while short saccades indicate a scanning behavior akin to ambient vision or more automatic (preattentive) processing.

Due to the design of the study, it is difficult to conclude whether the selective benefits of the info-vis display reflect limitations of color integration in displays for older adults, or the ways in which color was implemented in our info-vis condition. If the results reflect the latter, this could explain why significant improvements were only observed for a single attribute (gap coverage). The task improvement could be due to the substitution of textual data for the visual color scale. This change, unlike the other attributes, represents a transformation of the data from textual jargon, to a familiar color scale, which removes the need for knowledge of the specific meanings of each category. In the other three attributes, color was used in addition to the numeric and textual information (especially in the monthly premium and annual deductible attributes). Therefore, the transformation from textual information to the heat mapping color scale in the gap coverage attribute may be more beneficial for older adults than searching through data that is overlaid with color saturations which suggests rankings (that still contain numerical information that older adults could choose to use in their comparison rather than solely relying on the color saturations). Although the benefits of color info-vis are well established among younger adults, there may be limitations of benefits in older adults. Because older adults have the greatest amount of difficulty with information integration tasks [11], a configural display that illustrates relationships among different attributes using color and shape could have further boosted older adults' decision performance. Future research should examine how perceptual manipulations (eg, color, size, and shape) interact together and whether high difficulty comparisons and integration tasks can be simplified. This study did not examine the effects of size and color together or how these manipulations can improve specific types of data (eg, categorical vs interval). Another limitation of this study was the small sample size in each decision aid condition. As such, the results may provide low power to detect aid-related effects on performance.

Conclusions

Reducing the working memory demand of the task through the use of an ES can improve decision accuracy in certain cases. The results of this study indicate that color info-vis may be a viable ES for older decision makers for comparison tasks. Additionally, decision-making based on a single attribute can lead to better selection of drug plans for older adults. Instead of presenting all attributes at once for users to compare and contrast, each attribute could be presented individually and feature ES to reduce working memory demand. Faceted information retrieval is a filtering system that allows users to search along a specific feature or attribute [33]. Many search engines and retail websites utilize faceted search for improved navigation. This search method would yield a more manageable set of drug plans to choose from because less desirable plans would be filtered out with each attribute. Further research is needed to examine whether the use of color info-vis for each attribute could improve the quality of selected drug plans and reduce the time spent identifying the optimal choice compared to the current complex table design.

Acknowledgments

This research was partially funded by a Google Faculty Research Award.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Task analysis for choosing a prescription drug plan from the Medicare website.

[PDF File (Adobe PDF File), 54KB - humanfactors_v3i1e16_app1.pdf]

References

1. Czaja SJ, Sharit J, Nair SN. Usability of the Medicare health web site. JAMA 2008 Aug 20;300(7):790-792 [FREE Full text] [doi: 10.1001/jama.300.7.790-b] [Medline: 18714058]
2. Agree EM, King AC, Castro CM, Wiley A, Borzekowski Dina L G. "It's got to be on this page": Age and cognitive style in a study of online health information seeking. J Med Internet Res 2015;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
3. Hsu J, Fung V, Price M, Huang J, Brand R, Hui R, et al. Medicare beneficiaries' knowledge of Part D prescription drug program benefits and responses to drug costs. JAMA 2008 Apr 23;299(16):1929-1936. [doi: 10.1001/jama.299.16.1929] [Medline: 18430912]
4. Hanoch Y, Wood S, Barnes A, Liu PJ, Rice T. Choosing the right medicare prescription drug plan: The effect of age, strategy selection, and choice set size. Health Psychol 2011 Nov;30(6):719-727. [doi: 10.1037/a0023951] [Medline: 21604880]

5. Salthouse TA. Working memory as a processing resource in cognitive aging. Developmental Review 1990 Mar;10(1):101-124. [doi: 10.1016/0273-2297(90)90006-P]
6. McWilliams JM, Afendulis CC, McGuire TG, Landon BE. Complex Medicare advantage choices may overwhelm seniors--especially those with impaired decision making. Health Aff (Millwood) 2011 Sep;30(9):1786-1794 [FREE Full text] [doi: 10.1377/hlthaff.2011.0132] [Medline: 21852301]
7. Baddeley A. Working memory. Oxford: Oxford [Oxfordshire]: Clarendon Press; 1987.
8. Wickens C. Engineering psychology and human performance. New York, NY: Harper Collins Publishers; 1992.
9. Mata R, Schooler LJ, Rieskamp J. The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. Psychol Aging 2007 Dec;22(4):796-810. [doi: 10.1037/0882-7974.22.4.796] [Medline: 18179298]
10. Chen Y, Sun Y. Age differences in financial decision-making: Using simple heuristics. Educational Gerontology 2003 Aug;29(7):627-635. [doi: 10.1080/713844418]
11. Finucane ML, Slovic P, Hibbard JH, Peters E, Mertz CK, MacGregor DG. Aging and decision-making competence: An analysis of comprehension and consistency skills in older versus younger adults considering health-plan options. J. Behav. Decis. Making 2002 Apr;15(2):141-164. [doi: 10.1002/bdm.407]
12. Finucane ML, Mertz CK, Slovic P, Schmidt ES. Task complexity and older adults' decision-making competence. Psychol Aging 2005 Mar;20(1):71-84. [doi: 10.1037/0882-7974.20.1.71] [Medline: 15769215]
13. Tanius BE, Wood S, Hanoch Y, Rice T. Judgment and Decision Making. 2009 Feb. Aging and choice: Applications to Medicare Part D URL: http://journal.sjdm.org/81017/jdm81017.pdf [accessed 2016-05-19] [WebCite Cache ID 6hd0yzC27]
14. Miller GA. The magical number seven plus or minus two: Some limits on our capacity for processing information. Psychol Rev 1956 Mar;63(2):81-97. [Medline: 13310704]
15. Ratwani RM, Trafton JG, Boehm-Davis DA. Thinking graphically: Connecting vision and cognition during graph comprehension. J Exp Psychol Appl 2008 Mar;14(1):36-49. [doi: 10.1037/1076-898X.14.1.36] [Medline: 18377165]
16. Grierson HJ, Corney JR, Hatcher GD. Using visual representations for the searching and browsing of large, complex, multimedia data sets. International Journal of Information Management 2015 Apr;35(2):244-252. [doi: 10.1016/j.ijinfomgt.2014.12.003]
17. Pomerantz JR, Pristach EA. Emergent features, attention, and perceptual glue in visual form perception. J Exp Psychol Hum Percept Perform 1989 Nov;15(4):635-649. [Medline: 2531201]
18. Klix F, Hagendorf H. A functional account of age differences in memory. In: Human memory and cognitive capabilities: Mechanisms and performances: symposium in memoriam Hermann Ebbinghaus - 1885 - Berlin Humboldt University - 1985. Amsterdam: North-Holland; 1986.
19. Morrow DG, Rogers WA. Environmental support: An integrative framework. Hum Factors 2008 Aug;50(4):589-613. [Medline: 18767520]
20. Sharit J, Czaja SJ, Nair S, Lee CC. Effects of age, speech rate, and environmental support in using telephone voice menu systems. Hum Factors 2003;45(2):234-251. [Medline: 14529196]
21. Lohse GL. The role of working memory on graphical information processing. Behaviour & Information Technology 1997 Jan;16(6):297-308. [doi: 10.1080/014492997119707]
22. Pak R, Price MM, Thatcher J. Age-sensitive design of online health information: Comparative usability study. J Med Internet Res 2009;11(4):e45 [FREE Full text] [doi: 10.2196/jmir.1220] [Medline: 19917549]
23. Plude DJ, Doussard-Roosevelt JA. Aging, selective attention, and feature integration. Psychol Aging 1989 Mar;4(1):98-105. [Medline: 2803617]
24. Breslow LA, Ratwani RM, Trafton JG. Cognitive models of the influence of color scale on data visualization tasks. Hum Factors 2009 Jun;51(3):321-338. [Medline: 19750795]
25. Treisman A. Perceptual grouping and attention in visual search for features and for objects. J Exp Psychol Hum Percept Perform 1982 Apr;8(2):194-214. [Medline: 6461717]
26. Jacobson G, Gold M, Damico A, Neuman T, Casillas G. Medicare advantage 2016 data spotlight: Overview of plan changes. 2015 Dec 03. URL: http://kff.org/medicare/issue-brief/medicare-advantage-2016-data-spotlight-overview-of-plan-changes/ [accessed 2015-10-25] [WebCite Cache ID 6cXsLYqjf]
27. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A visualization tool to analyse usage of web-based interventions: The example of positive online weight reduction (POWeR). JMIR Human Factors 2015 May 19;2(1):e8. [doi: 10.2196/humanfactors.4310]
28. Galitz WO. The essential guide to user interface design: An introduction to GUI design principles and techniques. Indianapolis, IN: Wiley Technology; 2007.
29. Wickens CD, Carswell CM. The proximity compatibility principle: Its psychological foundation and relevance to display design. hum factors 1995 Sep 01;37(3):473-494. [doi: 10.1518/001872095779049408]
30. Payne JW. Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance 1976 Aug;16(2):366-387. [doi: 10.1016/0030-5073(76)90022-2]
31. Wechsler D. Wais-3 administration and scoring manual. USA: Psychological Corp; 1997.

32. Velichkovsky BM, Joos M, Helmert JR, Pannasch S. Two visual systems and their eye movements: Evidence from static and dynamic scene perception. Mahwah, NJ: Lawrence Erlbaum; 2005 Presented at: Conference of the Cognitive Science Society; 2005; Italy p. 2283-2288.
33. Tunkelang D. Faceted search. Synthesis lectures on information concepts, retrieval, and services 2009 Jan;1(1):1-80. [doi: 10.2200/S00190ED1V01Y200904ICR005]

Abbreviations

ANOVA: analysis of variance
color info-vis: color information visualization
ES: environmental supports
info-vis: information visualization
M: mean
s: seconds

Edited by G Eysenbach; submitted 04.09.15; peer-reviewed by D Morrow; comments to author 05.10.15; revised version received 16.11.15; accepted 04.01.16; published 01.06.16.

Please cite as:
Price MM, Crumley-Branyon JJ, Leidheiser WR, Pak R
Effects of Information Visualization on Older Adults' Decision-Making Performance in a Medicare Plan Selection Task: A Comparative Usability Study
JMIR Hum Factors 2016;3(1):e16
URL: http://humanfactors.jmir.org/2016/1/e16/
doi:10.2196/humanfactors.5106
PMID:27251110

Original Paper

University Students' Views on the Perceived Benefits and Drawbacks of Seeking Help for Mental Health Problems on the Internet: A Qualitative Study

Jade KY Chan[1], MSc; Louise M Farrer[2,3], PhD; Amelia Gulliver[2], PhD; Kylie Bennett[2], BA(Hons); Kathleen M Griffiths[2], PhD

[1]School of Psychology, Faculty of Science, University of New South Wales, Sydney, Australia

[2]National Institute for Mental Health Research, The Australian National University, Canberra, Australia

[3]Young and Well CRC, Melbourne, Australia

Corresponding Author:
Jade KY Chan, MSc
School of Psychology
Faculty of Science
University of New South Wales
Mathews Building
11 Botany Street
Sydney, NSW 2052
Australia
Phone: 61 2 6125 8859
Fax: 61 2 6125 8859
Email: chan.jadeky@gmail.com

Abstract

Background: University students experience high levels of mental health problems yet very few seek professional help. Web-based mental health interventions may be useful for the university student population. However, there are few published qualitative studies that have examined the perceived benefits and drawbacks of seeking help for mental health problems on the Internet from the perspective of university students.

Objective: To investigate the attitudes of university students on mental health help-seeking on the Internet.

Methods: A total of 19 university students aged 19-24 years participated in 1 of 4 focus groups to examine their views toward help-seeking for mental health problems on the Internet.

Results: Perceived concerns about Web-based help-seeking included privacy and confidentiality, difficulty communicating on the Internet, and the quality of Web-based resources. Potential benefits included anonymity/avoidance of stigma, and accessibility. Participants reported mixed views regarding the ability of people with similar mental health issues to interact on the Internet.

Conclusions: These factors should be considered in the development of Web-based mental health resources to increase acceptability and engagement from university students.

(JMIR Human Factors 2016;3(1):e3) doi:10.2196/humanfactors.4765

KEYWORDS

online; mental health; help seeking; university; students; qualitative

Introduction

Approximately 30% of university students meet criteria for a mental disorder [1,2]. The most commonly reported disorders in this population are substance use disorders, personality disorders, depression, anxiety disorders, and eating disorders [3]. Risk factors for developing a mental disorder among university students include experience of a stressful life event, living away from home, and experience of financial stress [2,3]. Despite the severity of the negative outcomes associated with untreated mental disorders, including poorer interpersonal relationships, lower engagement in campus activities, and increased risk of educational dropout [4], very few students receive appropriate mental health care [1]. Given the barriers

to seeking treatment reported by students (eg, fear of stigma, cost, and time constraints) [5,6], Web-based mental health interventions, such as peer-to-peer forums, screening tools, and educational and self-help programs, may be highly suited to the university population. Web-based interventions are easily accessible, can be utilized in private, are cost-effective, and typically require less time than face-to-face appointments [7]. Moreover, there is evidence that Web-based interventions targeting mental health are effective for university students [8] and that university students hold favorable attitudes towards Web-based mental health resources [9]. However, there is little exploration of concerns that university students may have regarding seeking help on the Internet, or why they may find Web-based mental health resources appealing. Qualitative research allows in-depth exploration of students' perceptions of these issues, which has key implications for the design of Web-based mental health resources targeting this population. The current study aims to investigate the attitudes of university students towards seeking mental health support on the Internet.

Methods

Design

A total of 4 focus groups were conducted with students from a large university in Canberra, Australia. Each focus group consisted of 4 or 5 participants (*n*=5, 5, 4, 5) with a total of 19 participants. Focus groups were chosen due to their potential to generate in-depth discussion. Given the general nature of the questions asked, discussion of personal experiences was not required. The duration of each session was approximately 1.5 hours. Ethics approval was granted by The Australian National University (ANU) Human Research Ethics Committee (2012/520). Cinema vouchers were offered to participants for their time and involvement in the focus groups.

Participants

Of the 19 university students who participated in the focus groups, 10 were women. The mean age of the sample was 21.6 years (range 19-24 years) and the mean duration of their university education was 3.1 years (range 1-5 years). Most participants were domestic students (*n*=12, 63%) from a range of disciplines (arts, law, commerce, engineering, science, music, and combined degrees from those faculties). Participants were recruited via email invitations distributed to students in residential colleges and students who had previously expressed interest in participating in mental health research.

Procedure

All focus groups were conducted by the primary facilitator (LF), a female postdoctoral research fellow and registered psychologist at the National Institute for Mental Health Research (NIMHR). During focus group meetings, 2 research assistants (AG and JC) were present to provide assistance and take field notes. The focus groups were audio recorded. On arrival, participants were provided with an information sheet, a consent form, and a short demographic questionnaire. At the beginning of each focus group session, the primary facilitator provided a brief introduction to the study and information about the purpose of the focus groups, confidentiality, and voluntary participation. Focus group participants were asked a series of questions relating to help-seeking preferences for mental health problems (both offline and on the Internet) and their views toward the development of a virtual mental health clinic for university students. Data relevant to the development of the virtual clinic have been published elsewhere (see [10] and [11]). The current paper focuses on participant responses to the following question: *"What do you think about using the Internet to get support for mental health or emotional problems?"* Participants who sought clarification about the meaning of "using the Internet" were provided with typical examples of Web-based mental health resources (eg, informational websites, self-help therapy programs, and peer-to-peer support networks such as forums and chat platforms).

Analysis Strategy

Data were analyzed using thematic analysis in NVivo 10 by the first author (JC). Participants' statements in response to the question above were coded based on a grounded theory approach [12], whereby similar concepts were grouped together into themes. During the discussion of Web-based help-seeking 2 major themes emerged: "concerns" and "benefits." Direct quotes are used to illustrate the themes and participants are identified by gender and a coded number (eg, F1 = female, participant number 1). The number of participants who endorsed each benefit and concern are also reported.

Results

Concerns Regarding Web-Based Mental Health Support

Participants reported a number of concerns about seeking help for mental health problems on the Internet including privacy and confidentiality (n=2), resource quality (n=3), and communication difficulties (n=4). A major concern was privacy and confidentiality:

> *I'd be worried about putting things in writing in the Internet. I don't know why. Like I know...that there'll have to be confidentiality around it, but the Internet just seems kind of insecure.* [F1]

Participants noted that while there is a wealth of mental health information on the Internet, it is scattered and difficult to find, which was viewed as a barrier to seeking help:

> *There's just so much, and that it's not really centralized and, you know, it's about knowing where to go as well.* [F2]

There was also skepticism about the quality and accuracy of information available on the Internet, as well as the ability for Web-based resources to address individualized problems:

> *I probably think it's a bit too generic for like, a personal problem.* [F3]

A final concern that emerged was that participants felt that it could be difficult to accurately portray emotions through writing on the Internet:

> *I think talking and typing it out is also different...It is a good idea but...maybe someone just can't get that*

across to someone else when you're just typing it out. [F4]

Benefits of Seeking Mental Health Support on the Internet

Participants reported several benefits of using Web-based resources to address mental health problems. These included anonymity/avoidance of stigma (n=3) and accessibility (n=3). The ability to seek help without other people knowing (and potentially passing judgement) was seen as particularly important:

> *The Internet is very useful and it's almost an entirely anonymous way to gather information so you can use it without, sort of revealing anything about yourself, about you personally...you can just take information without having to give any.* [M2]

Accessibility after business hours was also raised as an advantage of Web-based resources given that, in the participants' experience, mental health crises tended to occur at night.

> *It's readily accessible, so...if you're out of hours obviously you can go and see, like on a first-hand basis what...some advice is.* [F3]

Mixed Views

During a discussion of Web-based forums, participants reported mixed views regarding the ability for people with past or current experience of similar mental health problems to seek support from one another on the Internet. One participant expressed concern (n=1) that Web-based forums may compromise safety by enabling interaction between people who are distressed, thereby exacerbating their problems.

> *If you've got a bunch of really depressed people all together in a kind of, I guess, confined Internet space, they just make each other more depressed.* [M1]

For others (n=2), the ability of the Internet to connect people experiencing similar issues (eg, via forums) was viewed very positively:

> *I definitely feel it could be an important tool...I think there's no better way to help someone out than someone who's actually been through the same problems as you have.* [M3]

Discussion

Principal Findings

When asked directly about their views toward seeking Web-based support for mental health problems, participants raised a number of concerns. These related primarily to the privacy of Web-based resources, which has been noted as a key issue to consider in the development of health websites [13-16]. Additionally, the difficulty of communicating effectively on the Internet was raised. This echoes previous findings in a sample of university students [17], where students who preferred face-to-face therapy cited facilitation of communication as a primary reason for their preference. Additionally, difficulty with communication has previously been raised as a potential issue in Internet therapy [18]. However, Abbott and colleagues suggested this problem could be minimized by the user seeking clarification from the message source if there is a concern or misunderstanding [18].

Several participants mentioned the desire for centralized information and resources, suggesting that young people may feel overwhelmed when consulting the Internet for help with their mental health. Developers of Web-based mental health websites or interventions should aim to consolidate and present resources in a navigable and accessible way. Increasing awareness of websites with centralized information may also be beneficial to university students.

However, several positive attitudes toward Web-based help-seeking were raised, including the ability to access help privately and after hours. These are often cited as advantages of Web-based resources (eg, [7]). The Internet was also viewed as filling a gap for students who may wish to remain anonymous or avoid stigma, which is consistent with previous studies with other groups of young people [19]. The participants in the current study endorsed the idea of a Web-based forum for young people to relate to one another. Significantly, prior research has demonstrated that more than half of young people aged 18-25 years use forums for connecting with their peers, and for talking about their problems on the Internet [20]. However, moderation of Web-based interaction in forums is an important consideration for the development of a university virtual clinic, given the concern raised that forum discussion could exacerbate psychological distress. Participants also expressed a desire for relevant, centralized resources. Overall, a virtual clinic that offers tailored information, self-help programs, access to professional support, confidential screening and feedback, and peer support tools is likely to appeal to this population.

Limitations

There are several limitations to the present research that require consideration. Participants were not selected on the basis of current or previous experience with a mental health problem, and their mental health status was not assessed. It is unclear to what extent the views and intentions of students without mental health problems are applicable to those of students with a mental illness. However, several students chose to disclose either personal or close family members'/friends' experiences of a mental health problem. The views of all university students were considered valid for the purposes of the study, given that the aim of the virtual university clinic is to provide prevention services as well as treatment. Because participants self-selected to participate and this study was conducted in one university setting, the views expressed may not represent the views of all students within the participating university more broadly, or students from other universities. Finally, findings in the current paper are limited to the responses to a specific question of interest.

Implications

The findings of this research have implications for the development of Web-based mental health resources for university students. Despite their potential impact during emerging adulthood, mental health problems still remain undertreated in this vulnerable group, in part due to barriers to

help-seeking (eg, cost, time, accessibility). Web-based mental health resources have the potential to overcome these barriers; however, in order to optimize acceptability and engagement, it is critical that the concerns of university students are taken into account in the development of these resources in universities. Web-based resources should seek to offer security of information, anonymity, and treatment that is tailored to an individual's needs. Involving students in the intervention development process as co-designers may also address some of their concerns and improve eventual uptake of these services.

Acknowledgments

The authors wish to acknowledge Julia Reynolds, Anthony Bennett, Robert Tait, Philip Batterham, and Alison Calear for assistance with developing the focus group questions. This project was resourced by the Young and Well Cooperative Research Centre. The Young and Well CRC is established under the Australian Government's Cooperative Research Centres Program. KMG is supported by a National Health and Medical Research Council Fellowship (No. 1059620).

Conflicts of Interest

None declared.

References

1. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
2. Said D, Kypri K, Bowman J. Risk factors for mental disorder among university students in Australia: Findings from a web-based cross-sectional survey. Soc Psychiatry Psychiatr Epidemiol 2013 Jun;48(6):935-944. [doi: 10.1007/s00127-012-0574-x] [Medline: 22945366]
3. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu S, et al. Mental health of college students and their non-college-attending peers: Results from the National Epidemiologic Study on Alcohol and Related Conditions. Arch Gen Psychiatry 2008 Dec;65(12):1429-1437 [FREE Full text] [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
4. Salzer MS. A comparative study of campus experiences of college students with mental illnesses versus a general college sample. J Am Coll Health 2012 Jan;60(1):1-7. [doi: 10.1080/07448481.2011.552537] [Medline: 22171723]
5. Givens JL, Tjia J. Depressed medical students' use of mental health services and barriers to use. Acad Med 2002 Sep;77(9):918-921. [Medline: 12228091]
6. Mowbray CT, Megivern D, Mandiberg JM, Strauss S, Stein CH, Collins K, et al. Campus mental health services: Recommendations for change. Am J Orthopsychiatry 2006 Apr;76(2):226-237. [doi: 10.1037/0002-9432.76.2.226] [Medline: 16719642]
7. Griffiths KM, Christensen H. Internet-based mental health programs: A powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
8. Farrer L, Gulliver A, Chan Jade K Y, Batterham PJ, Reynolds J, Calear A, et al. Technology-based interventions for mental health in tertiary students: Systematic review. J Med Internet Res 2013 May;15(5):e101 [FREE Full text] [doi: 10.2196/jmir.2639] [Medline: 23711740]
9. Ryan ML, Shochet IM, Stallman HM. Universal online interventions might engage psychologically distressed university students who are unlikely to seek formal help. Adv Ment Health 2014 Dec 17;9(1):73-83. [doi: 10.5172/jamh.9.1.73]
10. Farrer L, Gulliver A, Chan JK, Bennett K, Griffiths KM. A virtual mental health clinic for university students: A qualitative study of end-user service needs and priorities. JMIR Ment Health 2015;2(1):e2 [FREE Full text] [doi: 10.2196/mental.3890] [Medline: 26543908]
11. Gulliver A, Bennett K, Bennett A, Farrer LM, Reynolds J, Griffiths KM. Privacy issues in the development of a virtual mental health clinic for university students: A qualitative study. JMIR Ment Health 2015;2(1):e9 [FREE Full text] [doi: 10.2196/mental.4294] [Medline: 26543915]
12. Martin P. Grounded theory and organizational research. J Appl Behav Sci 1986 Apr 01;22(2):141-157. [doi: 10.1177/002188638602200207]
13. D'Andrea G. Health web site accreditation: Opportunities and challenges. Manag Care Q 2002;10(1):1-6. [Medline: 15988948]
14. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, Chumbler NR. Lessons learned from usability testing of the VA's personal health record. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i13-i17 [FREE Full text] [doi: 10.1136/amiajnl-2010-000082] [Medline: 21984604]
15. Kitayama K, Stockwell MS, Vawdrey DK, Peña O, Catallozzi M. Parent perspectives on the design of a personal online pediatric immunization record. Clin Pediatr (Phila) 2014 Mar;53(3):238-242. [doi: 10.1177/0009922813506608] [Medline: 24137033]
16. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online access to doctors' notes: Patient concerns about privacy. J Med Internet Res 2013;15(9):e208 [FREE Full text] [doi: 10.2196/jmir.2670] [Medline: 24072335]

17. Horgan A, Sweeney J. Young students' use of the Internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]
18. Abbott JM, Klein B, Ciechomski L. Best practices in online therapy. J Technol Hum Serv 2008 Jul 03;26(2-4):360-375. [doi: 10.1080/15228830802097257]
19. Gulliver A, Griffiths KM, Christensen H. Barriers and facilitators to mental health help-seeking for young elite athletes: A qualitative study. BMC Psychiatry 2012;12:157 [FREE Full text] [doi: 10.1186/1471-244X-12-157] [Medline: 23009161]
20. Ellis LA, Collin P, Hurley PJ, Davenport TA, Burns JM, Hickie IB. Young men's attitudes and behaviour in relation to mental health and technology: Implications for the development of online mental health services. BMC Psychiatry 2013;13:119 [FREE Full text] [doi: 10.1186/1471-244X-13-119] [Medline: 23601273]

Edited by G Eysenbach; submitted 29.05.15; peer-reviewed by N Kanuri, J Saleem, K Vallury, A Horgan, EB Davies; comments to author 29.06.15; revised version received 30.07.15; accepted 22.09.15; published 19.01.16.

Please cite as:
Chan JKY, Farrer LM, Gulliver A, Bennett K, Griffiths KM
University Students' Views on the Perceived Benefits and Drawbacks of Seeking Help for Mental Health Problems on the Internet: A Qualitative Study
JMIR Human Factors 2016;3(1):e3
URL: http://humanfactors.jmir.org/2016/1/e3/
doi:10.2196/humanfactors.4765
PMID:27026140

Original Paper

A Mixed-Methods Approach to the Development, Refinement, and Pilot Testing of Social Networks for Improving Healthy Behaviors

Sarah Hales[1], PhD; Gabrielle Turner-McGrievy[1], PhD; Arjang Fahim[2], PhD; Andrew Freix[2]; Sara Wilcox[3], PhD; Rachel E Davis[1], PhD; Michael Huhns[2], PhD; Homayoun Valafar[2], PhD

[1]Arnold School of Public Health, Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, United States

[2]Department of Computer Science, University of South Carolina, Columbia, SC, United States

[3]Arnold School of Public Health, Department of Exercise Science, University of South Carolina, Columbia, SC, United States

Corresponding Author:
Sarah Hales, PhD
Arnold School of Public Health
Department of Health Promotion, Education, and Behavior
University of South Carolina
Discovery I Building
915 Greene St.
Columbia, SC,
United States
Phone: 1 843 670 5317
Fax: 1 803 777 6290
Email: bridges5@mailbox.sc.edu

Abstract

Background: Mobile health (mHealth) has shown promise as a way to deliver weight loss interventions, including connecting users for social support.

Objective: To develop, refine, and pilot test the Social Pounds Off Digitally (POD) Android app for personalized health monitoring and interaction.

Methods: Adults who were overweight and obese with Android smartphones (BMI 25-49.9 kg/m^2; N=9) were recruited for a 2-month weight loss pilot intervention and iterative usability testing of the Social POD app. The app prompted participants via notification to track daily weight, diet, and physical activity behaviors. Participants received the content of the behavioral weight loss intervention via podcast. In order to re-engage infrequent users (did not use the app within the previous 48 hours), the app prompted frequent users to select 1 of 3 messages to send to infrequent users targeting the behavioral theory constructs social support, self-efficacy, or negative outcome expectations. Body weight, dietary intake (2 24-hr recalls), and reported calories expended during physical activity were assessed at baseline and 2 months. All participants attended 1 of 2 focus groups to provide feedback on use of the app.

Results: Participants lost a mean of 0.94 kg (SD 2.22, *P*=.24) and consumed significantly fewer kcals postintervention (1570 kcal/day, SD 508) as compared to baseline (2384 kcal/day, SD 993, *P*=.01). Participants reported expending a mean of 171 kcal/day (SD 153) during intentional physical activity following the intervention as compared to 138 kcal/day (SD 139) at baseline, yet this was not a statistically significant difference (*P*=.57). There was not a statistically significant correlation found between total app entries and percent weight loss over the course of the intervention (*r*=.49, *P*=.19). Mean number of app entries was 77.2 (SD 73.8) per person with a range of 0 to 219. Messages targeting social support were selected most often (32/68, 47%), followed by self-efficacy (28/68, 41%), and negative outcome expectations (8/68, 12%). Themes from the focus groups included functionality issues, revisions to the messaging system, and the addition of a point system with rewards for achieving goals.

Conclusions: The Social POD app provides an innovative way to re-engage infrequent users by encouraging frequent users to provide social support. Although more time is needed for development, this mHealth intervention can be disseminated broadly for many years and to many individuals without the need for additional intensive in-person hours.

(JMIR Human Factors 2016;3(1):e8) doi:10.2196/humanfactors.4512

KEYWORDS

mHealth; obesity; weight loss; social support; social cognitive theory

Introduction

Rates of overweight and obese US adults remain high with 69% having a body mass index (BMI) greater than 25 kg/m^2 [1]. Obesity and being overweight are associated with type 2 diabetes, cardiovascular disease, arthritis, asthma [2,3], and some cancers—thyroid [4], colon, breast (in postmenopausal women), endometrium, esophagus, and kidney [5]. Behavioral interventions that target improvements in diet and physical activity are effective ways to help people lose weight and decrease chronic disease risk factors [6].

The use of apps on mobile devices (eg, smartphones and tablets) has the potential to improve how individuals monitor health behaviors by serving as a convenient platform to connect users to one another via online social networks. Mobile health (mHealth) holds promise as an effective method of delivering behavioral interventions addressing diet and physical activity, and it is less time-intensive than in-person, individual, or group sessions [7]. Mobile phone ownership is pervasive; 85% of US adults report owning a mobile phone with half owning smartphones [8]. Smartphone ownership cuts across ethnic groups with 49% of Hispanics, 47% of African Americans, and 42% of whites owning smartphones [8]. While there has been emerging research in the area of using mHealth technologies to help people achieve a healthy weight, few studies have used an entirely mobile device-based approach to deliver a behavioral weight loss intervention. Furthermore, many of the mobile-based weight loss apps available (both free and paid) do not include the evidence-based techniques used in traditional (ie, clinic-based) weight loss interventions [9,10].

Weight loss programs have been developed and delivered via Internet and other Web-based platforms as well as through social media (eg, Facebook [11] and Twitter [12]) to promote weight loss and reduce health risks of chronic disease. Frequent participant engagement with social media in the context of these weight loss interventions has been shown to be related to weight loss [11-13]. While there are many benefits to delivering weight loss interventions using remote methodologies, maintaining participant engagement over time and providing sufficient social support in these types of interventions can be a challenge [14].

The overarching objective of this line of research has been to design an app that can be used to monitor and test scientific hypotheses related to optimal matching of participants to provide support for collective weight loss in the context of mobile interventions. The primary goal for this pilot study was to solicit participant feedback to refine the Social Pounds Off Digitally (Social POD) app for use in a larger pilot randomized clinical trial (RCT). The Social POD app was developed by a transdisciplinary team of researchers including experts in health behavior, nutrition, computer science, psychology, exercise science, and social work. The analysis sought to answer the following questions: (1) What features of the Social POD app needed to be refined or developed to further incentivize participants to use the app? (2) Were there any significant changes from pre- to poststudy in participant weight, calories consumed, and reported intentional physical activity? (3) Was participant weight loss correlated with frequency of app use over the course of the study?

Methods

Participant Recruitment

Men and women who were overweight or obese with Android smartphones (BMI 25-49.9 kg/m^2; N=9) were recruited in South Carolina for a 2-month weight loss pilot intervention to test usability and provide feedback to be used in the refinement of the smartphone app. Participants were recruited via worksite listserv announcements, flyers, and newspaper advertisements. Exclusion criteria included not having an Android phone, previous stroke or heart attack diagnosis, diagnosis of diabetes and using insulin or oral medications to control diabetes, BMI outside the range of 25.0-49.9 kg/m^2, unable to attend required meetings, unable to access the Internet using a computer for completing assessments, having a psychiatric illness, receiving treatment for drug or alcohol dependency, having an eating disorder, participating in another weight loss program, being pregnant or planning on becoming pregnant during the study, and breastfeeding. Participants were excluded for endorsing any of the first 4 items on the revised Physical Activity Readiness Questionnaire (PAR-Q) [15,16]: (1) informed by a doctor that they have a heart condition and should only participate in physical activities approved by a doctor, (2) feeling chest pain when engaging in physical activity, (3) experiencing chest pain in the past month when not engaging in physical activity, and (4) ever losing their balance and becoming dizzy or ever losing consciousness. If participants reported a bone or joint problem that could be made worse by participating in physical activity (item 5 of the PAR-Q) or taking blood pressure medication (item 6 of the PAR-Q), they were required to have a physician consent form completed to participate in the study. Participants received US $30 for completion of assessments at baseline, US $15 after the 1-month focus group, and US $15 after the 2-month follow-up weight assessment.

Intervention Implementation

Participants attended 4 in-person meetings: an orientation session to learn about the study and complete baseline dietary assessments; a training session to learn how to download and use the Social POD app and podcasts and to collect baseline height and weight measurements; a mid-study focus group at 1 month to provide feedback regarding the usability of the Social POD app, provide suggestions for improvement, and collect 1-month weight measurements; and an end-of-study session to provide 2-month weight measurements. All participants provided written consent. This study was approved by the University of South Carolina institutional review board.

Participants were instructed to track their total calories from all meals, snacks, and beverages consumed; minutes of intentional physical activity; and body weight each day for the duration of

the 2-month usability study. Participants were instructed to use MyFitnessPal or LoseIt, free commercial diet-tracking apps with extensive nutrient databases, to look up calorie information for food and beverages consumed. Participants were then asked to transfer total calories from each meal and snack consumed for the day to the Social POD app. Screenshots of the tracking features are shown in Figure 1.

Participants received within-app notifications at certain times throughout the day to remind them to self-monitor (promoting self-regulation) diet, minutes of physical activity, and total body weight each day. Participants could view a history of all dietary, activity, and weight information entered on the within-app history screen. Participants could view weight entered on a weight graph. Participants who were frequently using the Social POD app (users who entered information in the app in the past 48 hours) were prompted via notifications to send encouraging messages to other group members who had not entered data in the app over the previous 48 hours (infrequent users). Messages were sent from frequent users by clicking the notification to send a message to an infrequent user, selecting one of three options listed on the message selection screen, and clicking Send. Screenshots of the message log, message selection, and history screen of the Social POD app features are shown in Figure 2.

This study is novel in that it used theory-based messages designed by the researchers to help re-engage infrequent participants over the course of the 2-month intervention. The participants were matched to provide support to one another based on principles of recommender systems used by some websites and applications (eg, Amazon and Netflix) [17-20], which filter information to match users based on user history or preferences [21]. Frequent users were randomly matched to provide support (by sending a theory-based message) to help re-engage infrequent users in this intervention. Social Cognitive Theory (SCT) [22,23] was used as a framework to design user-to-user messages that targeted social support [13], self-efficacy [22,23], and outcome expectations [24] regarding self-monitoring behavior (ie, targeting self-regulation) [23] of diet, physical activity, and weight. SCT is the belief in the reciprocal relationship between cognitions, environmental influences, and behavior [22,25]. Constructs from SCT were selected for this study based on results from a previous Internet-based weight loss intervention, which found that targeting these specific constructs led to healthier diet and physical activity behaviors and resulted in a reduction of body weight among study participants [26]. The interventionists created messages for each of the three social construct theories and prompted frequent users to select 1 of 3 messages to send to an infrequent user. Table 1 provides examples of each message type as well as the SCT construct targeted.

Table 1. User-to-user message types by social cognitive theory construct targeted.

SCT construct targeted	Construct definition	Sample message
Self-efficacy	One's belief in the ability to perform specific tasks and overcome barriers.	"I found some light recipes on the Internet and they look pretty good. Nutrition info is listed, so they're easy to track too."
Social support	Support from others, which can take many forms including information, suggestions, or advice.	"Haven't seen you log anything in the app lately. We miss you!"
Outcome expectations	Expected outcomes of behaviors.	"I've never really succeeded at a diet before, but I think tracking my calories, weight, and exercise has to help this time around!"

Participants were provided with 3 20-minute, evidence-based weight loss podcasts each week. Podcasts were uploaded to the study website, and participants were sent a reminder that a new podcast was available every Monday, Wednesday, and Friday via email. Participants were instructed to listen to the 3 podcasts within the week but could listen at a time and place of their choosing. Podcasts were informed by SCT and provided participants with a range of weight loss topics. Podcast topics included nutrition and exercise information, an audio diary tracking the weight loss progress with challenges experienced by a male and female character, and a weight loss soap opera depicting the challenges of overcoming social barriers to weight loss, with a goal-setting activity related to weight loss at the end of each episode. Specific information regarding the development and testing of these podcasts in previous interventions can be found elsewhere [12,27].

Figure 1. Screenshots (left to right) of the home, meal tracking, and physical activity tracking screens on the Social POD app.

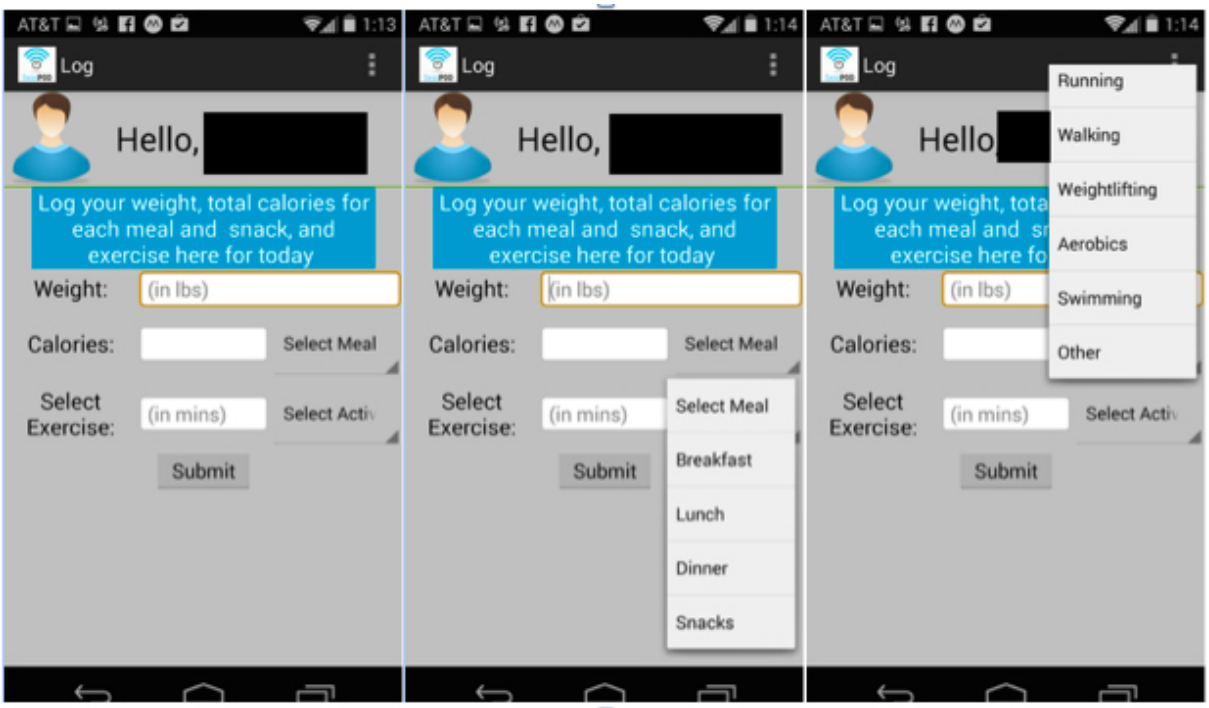
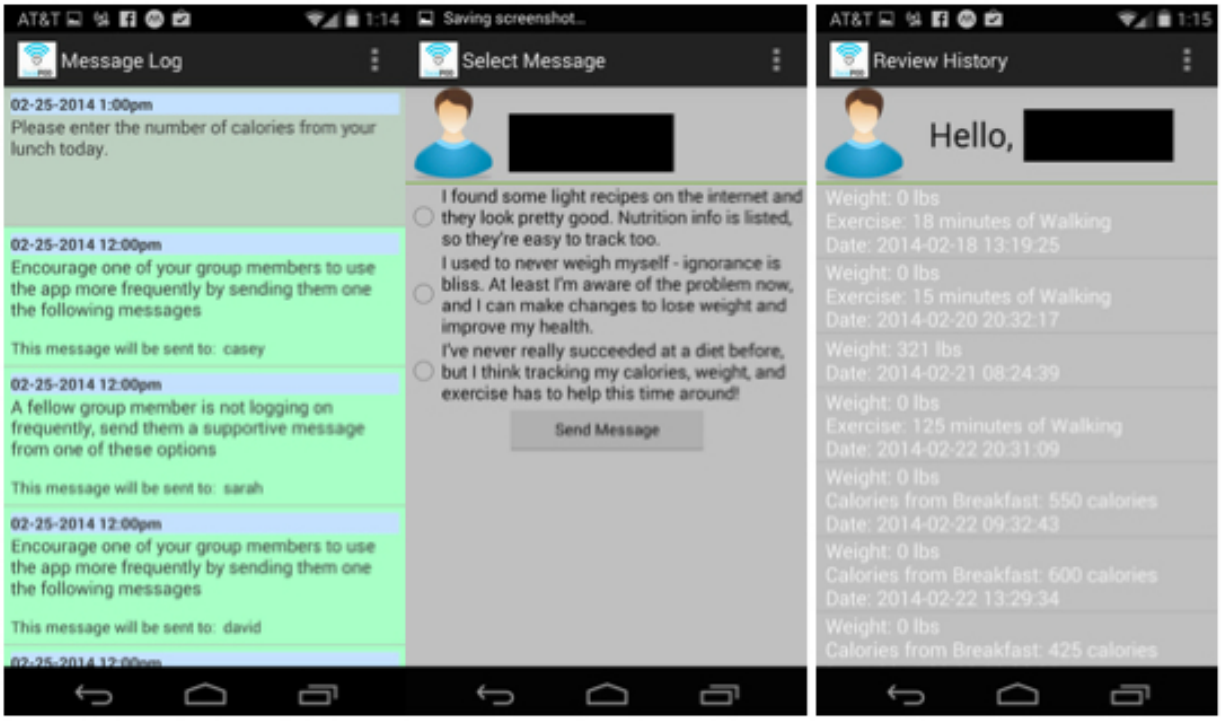


Figure 2. Screenshots (left to right) of the message log, message selection, and history screens on the Social POD app.



Outcome Measures

Participants completed online questionnaires assessing demographic characteristics. Height was measured at baseline assessment only using a Seca 213 (Hamburg, Germany) calibrated stadiometer. Weight was measured at baseline and 2 months using a Seca 869 calibrated digital scale accurate to 0.01 kg. Dietary measures were completed by participants using the National Cancer Institute's Automated Self-Administered 24-Hour Dietary Recall (ASA24) [28]. Each participant completed the ASA24 online for 1 unannounced weekday and weekend day at baseline and 2 months. The following previously validated questionnaires were completed by participants at baseline and 2 months unless otherwise specified: the Paffenbarger Physical Activity Questionnaire to determine calories expended during activity [29], the 20-item Weight Efficacy-Lifestyle Questionnaire to measure participant self-efficacy [30], the 44-item Big Five Inventory (baseline only) as a measure of personality characteristics [31], and the Sallis Social Support Scale [32] and Physical Activity Social

Influence Scale as measures of social support from family and friends (modified to include online social networks) [33]. Objective measures of frequency and duration of app use were collected unobtrusively by a secure network server following the intervention (at 2 months). All baseline measurements were collected prior to the administration of the weight loss intervention.

Participants attended 1 of 2 focus group sessions at 1 month to provide feedback regarding their use of the intervention components (Social POD app and podcasts). The focus groups were conducted by a trained focus group facilitator, and questions were asked using a semistructured interview guide developed by the investigators that included prescripted and spontaneous probes designed to elicit information from participants regarding their experience with the Social POD app. Questions were designed to assess participant satisfaction and dissatisfaction with the weight loss podcasts; the Social POD app including the tracking, prompting, and messaging features; and the weight graph and history features for future revisions. Participants were also asked if they would like incentives for using the app, and if so, how they would like an incentive system to be structured. Information regarding aesthetics and appearance of the app was solicited as well.

Focus group sessions were taped using an audio recorder, and all participants were instructed to use their study identification number to protect confidentiality during the focus groups. A trained graduate research assistant was present at the focus group sessions to take field notes during the sessions. The focus group facilitator completed detailed memos following each midstudy focus group session to document salient themes from participant discussions. Recordings from the 1-month focus group were transcribed verbatim by the research assistant and were cross-referenced with recordings for accuracy.

During the study (weeks 1 through 7) participants completed brief weekly online surveys each Friday to report the number of podcasts listened to and the number of days they used the Social POD app that week. Participants also reported any problems experienced using the Social POD app during the past 7 days and any suggestions they had to improve the app. If participants reported problems or suggestions for the Social POD app, they were then asked to write a detailed explanation.

Statistical Analysis

All qualitative data were collected and analyzed between February and August 2014 using NVivo for Mac Beta 2014 statistical software (QSR International) to extract a combination of a priori and emergent themes from the focus groups to guide the revision of the Social POD app. Open-ended responses from weekly surveys regarding problems experienced and suggestions for improvements to the Social POD app were coded as well. Stata version 13.1 (StataCorp) software was used to detect statistical significance for all quantitative data analysis. Paired samples *t* test was used to test for statistically significant differences from pre- to posttest for weight, calories consumed, and calories expended during reported physical activity. The effect size for weight loss was also calculated. Correlation between total app entries and percent weight loss at 2 months was examined. Correlation between total days of objectively measured app use and total days of self-reported app use (via weekly questionnaires) by participants was examined. Fisher exact test was used, due to the small sample size, to detect significant differences in user-to-user message selection among participants categorized into subgroups based on amount of weight loss (dichotomized into groups of high versus low weight loss at the median split of percent weight loss). The sample size for this study was determined for qualitative analysis based on expert recommendations for mHealth usability testing studies [34,35]; therefore, this study was not powered to detect significant changes in within-group weight. Changes in weight and weight-related behaviors, however, are presented.

Results

Demographic Characteristics

A total of 34 individuals inquired about the study, 15 individuals qualified to participate, and 9 completed baseline measures and received the intervention. Major reasons why participants did not qualify included not having an Android phone (9/19, 47%) and meeting medical exclusion criteria (3/19, 16%). Most participants (7/9, 78%) attended the 1-month focus group sessions (2 infrequent app users did not attend), and all participants completed the poststudy questionnaire and unannounced dietary recalls and had their body weight measured poststudy. The response rate for weekly surveys was 62% (39/63). Baseline demographics of the study sample are presented in Table 2.

Table 2. Baseline demographic characteristics.

Characteristics		n=9
Age (years), mean (SD)		39 (14.5)
Gender, n (%)		
	Female	8 (89)
	Male	1 (11)
Race, n (%)		
	Black	3 (33)
	White	6 (67)
Weight (kg), mean (SD)		91.5 (19.1)
Body mass index, mean (SD)		35.5 (7.1)
Educational attainment, n (%)		
	Advanced	3 (33)
	College	5 (56)
	Some college	1 (11)
Marital status, n (%)		
	Divorced or separated	1 (11)
	Married	2 (22)
	Single	4 (44)
	Other	2 (22)

Focus Group Themes

Major themes that emerged during focus group discussions are reported below with illustrative quotes from participants. These themes included Social POD app functionality, improvements to the Social POD app and podcasts, additions to the Social POD app, incentive and goal system for the Social POD app, and satisfaction with the Social POD app and podcasts.

Functionality of Social POD App

Functionality problems reported by participants included issues with the notification system that prompted participants to track their daily diet, weight, and minutes of physical activity. Reported functionality issues with the notification system mentioned during the focus groups as well as on the weekly surveys included 1 failure to receive notifications and 1 failure of notifications to connect with the correct screen within the Social POD app. The same participant reported 2 instances of spontaneous crashing of the Social POD app on the weekly surveys.

Improvements to Social POD App and Podcasts

Themes relating to suggested improvements for the Social POD app were varied and included adding additional features and modifying existing features. Modifying the color scheme and integrating options for personalization were suggested. One participant suggested using colors that “get your attention more” by replacing the grey and dark blue used in the current version of the Social POD app with brighter colors. Another participant mentioned she would like to have had the opportunity to choose an avatar or icon for her home screen to customize the appearance of her app. Having the ability to track calories consumed on previous dates was another requested modification. Most participants (3 out of 4 participants in this group) reported that the notifications to track their diet, physical activity, and weight were helpful as reminders to track their healthy behaviors, although 1 participant said she found them to be annoying.

There were not many suggested improvements for the podcasts during the focus groups, but participants did note problems with sound quality and variation in the volume level. A participant in the focus groups mentioned that there was a difference in the volume level of the various segments of the podcasts, specifically with the soap opera portion. Participants recommended leveling the volume and improving the sound quality of the podcast episodes prior to future use. While all participants reported that they valued the information provided in the podcasts and felt satisfied with the various segments, some participants (n=3) reported they did not like the soap opera portion of the podcast episodes. A participant shared, “I don’t particularly like the soap opera,” and 2 members of the group agreed with this statement and said this story line was “geared toward young dating women” and “negative.” Another participant would have enjoyed the soap opera portion of the podcast more if all of the episodes could be listened to consecutively rather than having to wait until the next episode was available to download.

Additions to the Social POD App

Suggested additions to the Social POD app included adding a database of common foods and beverages within the app (versus having to use another app to look up calorie information) and incorporating an incentive system. Increasing the amount of

praise provided by the study staff for entering weight and achieving weight loss–related goals was a highly suggested addition to the app. All of the participants that attended the focus groups (n=7) expressed their desire to see how others were doing in the program and to send other participants encouragement for achieving weight loss goals. A participant mentioned it would be interesting to see if messages sent to other users actually motivated them to use the app more often.

> *I think it'd be interesting to hear somebody from the receiving end to see if that motivated them to do something.*

Incentive and Goal System of the Social POD App

All participants in the focus groups reported they would like an incentive system with rewards for using the tracking components of the Social POD app. Participants suggested that points should be redeemed for prizes.

> *I'm all about a token economy, so yes,...I think like a five year old, so yeah, that [prizes] really does motivate me.*

Participants recommended earning points specifically for completing physical activity.

> *I haven't been nearly as active as I would have liked to have been...having a point reward system for that would be an incentive for me to move more.*

A participant mentioned that creating a competition among study participants or giving prizes for personal bests would be motivating.

> *I certainly like prizes [laughter]. It could also be fun to have a competition between the Social POD users so you could have the leader in points receive some prize or even personal bests, like if you have a 5-week program and your best week you could get some rewards.*

Participants recommended the integration of preset or tailored goals to limit the possibility that some might set unrealistic goals for themselves and give up on the program.

> *First participant: Think of the other side if you decide for yourself some goals, and for some reason you cannot make them, how would you feel?*
>
> *Second participant: Really terrible, but if it was somehow like...If the system could somehow help you set reasonable goals instead of like high in the sky things that I know I'm not going to do. People do tend to set lofty goals for themselves when really it should be small, measurable, incremental changes at first.*

Satisfaction With Social POD App and Podcasts

Themes relating to satisfaction with Social POD app components included the user-to-user messaging system. Some participants preferred the prewritten user-to-user messages because they seemed more professional than allowing participants to create their own user-to-user messages.

> *I think the prepopulated messages is [are] the best because it's professional.*

Overall, participants reported satisfaction with the messaging system. A participant reported that she would have been less likely to send messages to others if she had to write the messages herself.

> *I know that I would be less likely to send a message to someone if I had to write it myself...so having some [messages] to choose from where it takes just a second to do, I'm more likely to do that.*

Overall, participants reported the Social POD app was simple and easy to use. Participants reported they liked the convenience and ability to use their smartphones to track their diet, which they found to be more inconspicuous than other methods (eg, using a calorie book).

> *Every now and then you feel like James Bond because if you are at the table an[d] you still have your cellphone with you an[d] you turn on the app and...nobody knows what you do...an[d] you just...key in everything an[d] done. I mean [at] the end of the day...I've done something for myself today.*

Participants also mentioned that the app gave them the motivation they needed to change their diet and physical activity behaviors to lose weight.

> *It's definitely been a useful exercise...my thing is process, not perfection...in my case I'm doing a lot more than I was doing before.*

All participants in the focus groups volunteered that they valued the information provided in the podcasts.

> *I liked the information...it reinforces what I'm sure most of us already know.*

Participants reported a variety of methods used to listen to the weekly podcasts. A participant created an icon on the phone to easily access the podcasts and even set reminders to listen to the podcasts throughout the week.

> *I've set up the podcasts as an icon on my phone. So then I just go to the website and can play it from there, and then I have an alarm set up for Monday, Wednesday, and Friday because otherwise I'll forget to [listen] even with an email I'll probably forget.*

Another participant reported listening to the podcasts from a laptop.

> *It's easy, I told you, I listen to this on my laptop, and once I get the email, I just click on the link on my media system in the laptop [and] just open up, and I can see everything there. I wish they could stay. I wished at the end of the study [the podcasts] would not go away.*

Participants reported listening to the podcasts in a variety of settings. Some participants reported listening to the podcasts while in the car.

> *I like to listen to it in the car when I'm going to be in the car for twenty or thirty minutes.*
>
> *I like to listen to them in the car or when I'm getting ready in the morning and that's a good time to do it, so I appreciate that your assistant sends them early*

> *in the morning, so I have a good time, a whole day really to listen to it.*

All focus group participants mentioned that they liked at least one of the various segments included in the podcast episodes.

> *I also really like the section where you listen to someone's diary...that's helpful.*

Another participant liked the diversity of characters included in the podcasts.

> *I like there's an array of people who share. Older gentlemen, the women, whoever you seem to identify with you can find somebody who is good for you.*

Suggested Improvements

On the weekly surveys, participants suggested many of the same improvements discussed in the focus group sessions. Suggestions for improvements to the Social POD app included adding more vibrant colors to the app, adding customization such as a personal avatar, integrating a nutrient database, adding the ability to track diet and activity for previous dates, and providing more praise for losing weight and/or increasing physical activity. Additional suggestions made on the weekly surveys not mentioned in the focus groups included adding additional activity options for tracking physical activity (n=1), giving participants the ability to customize the user-to-user messages (n=1), and revising the user-to-user messages to sound more motivating and encouraging (n=1).

Planned Revisions of the Social POD App and Podcasts Based on Focus Group Themes

A news feed will be developed for the Social POD app for participants to view the progress of other users with weight loss-related goals. Participants will be able to send others encouragement for achieving these goals (eg, logging 30 minutes of exercise, logging diet and weight) through the news feed, targeting positive reinforcement [22]. Other revisions to the next iteration of the Social POD app will include the ability to earn points on a Point Tracker within the app for self-monitoring diet, physical activity, and weight and for sending others encouragement through the news feed. Points will be redeemed for study-provided prizes at the end of the pilot RCT, targeting reinforcement [22]. A weight loss competition among participants, as suggested by one participant, will not be integrated into the revised version of the Social POD app to minimize the potential risk of harm to some participants who are not achieving weight loss goals as quickly as others. Integrating a database of food and beverages similar to commercial diet tracking apps to view the nutrient content of commonly consumed items was recommended and will be incorporated into the calorie tracking features of the Social POD app to facilitate self-regulation [23]. A more extensive list of activities will be incorporated to the physical activity tracker in the revised Social POD app, also promoting self-regulation [23]. The color scheme of the app will be modified to include colors that are brighter and more eye-catching, and the option to add an avatar on the home screen will be incorporated. User-to-user messages will be revised to include more encouraging statements to better re-engage infrequent users with the Social POD app. Planned revisions for the podcasts include rerecording the podcast episodes to improve the sound quality and volume across segments within the episodes.

Quantitative Results

Participants lost a mean of 0.94 kg (SD 2.22). Differences in mean participant weight before (91.48 kg, SD 19.08; 95% CI 76.82-106.15) and after (90.55 kg, SD 20.01; 95% CI 75.17-105.93) the 2-month intervention were not statistically significant (P=.24, d=.05, r=.02). Participants reported expending a mean of 171 kcal/day (SD 153) during intentional physical activity following the intervention as compared to 138 kcal/day (SD 139) at baseline, yet this was not a statistically significant difference (P=.57). Participants reported consuming significantly fewer calories following the intervention (1570 kcal/day, SD 508) than before (2384 kcal/day, SD 993, P=.01).

Mean number of Social POD app entries over the course of the 2-month usability study was 77.2 (SD 73.8, 95% CI 17.66-133.68) with a minimum of 0 and a maximum of 219 entries. Messages were sent by frequent users targeting social support most often (32/68, 47%), followed by self-efficacy (28/68, 41%) and outcome expectations (8/68, 12%). There was not a statistically significant correlation between total app entries and percent weight loss over the course of the intervention (r=.49, P=.19). There was no difference in the type of message selected (self-efficacy, social support, and outcome expectation) between those participants who were successful at weight loss as compared to those who were less successful (defined at median split in percent weight loss) (P=.79).

On the weekly surveys, participants reported listening to an average of 2.24 podcast episodes (SD 1.50; minimum 0, maximum 6) per week, and they reported using the Social POD app an average of 4.5 days (SD 2.25; minimum 0, maximum 7). There were 7 reports of problems using the Social POD app by 3 participants over the course of the 2-month study. Reported problems from weekly surveys included spontaneous crashing (2 times) and notifications linking with the incorrect screen in the Social POD app (1 time); there were no explanations for the other reported problems on the weekly surveys.

There were 2 participants with no objective measure of Social POD app use over the course of the study. A participant reported using the Social POD app 1 day during the study, and she specified on the weekly survey that she did not use the app because she did not have enough time. The other participant without objective app use data reported on the weekly surveys that she used the app a total of 21 days during the study. The total number of days participants used the app (as objectively measured by the app) and total number of days of self-reported app use by participants via weekly surveys was highly correlated (r=.87, P<.01), indicating that self-report was a reliable measure of app engagement in this study.

Discussion

Comparison to Prior Work

While there has been much work in the area of mobile app development for health, there is currently little published research in the area of development and testing of new mobile apps for weight loss among adults who are overweight and

obese. There are several recent studies documenting the development and testing of new apps for weight loss in adolescents [36], apps for increasing physical activity and reducing screen time among adolescent males [37], and apps for predicting the risk of childhood obesity among infants [38]. Several recent studies report on the development and testing of apps for modification of diet and physical activity among the young adult population [39-41].

In a similar mixed-methods usability study, Morrison and colleagues used a computer-based weight loss program in conjunction with an Android mobile app to improve participant goal setting and motivation to achieve weight management goals among young adult participants over a 4-week period [39]. This app also offered participants the opportunity to set notifications, similar to those in the Social POD app, as reminders to use the diet and physical activity diary features. Participants could choose when and if they would like to receive notifications in an effort to improve the usability of and satisfaction with the mobile app [39]. It was found that only about half of the participants used the self-monitoring features of the app over the course of this usability testing [39]. While 1 participant in the Social POD app study noted that the notifications could be "annoying," others found the notifications, which were preset by study coordinators, a helpful reminder to self-monitor their behaviors and weight and said that they otherwise would not have performed this task. Prior research has also demonstrated that weight loss is improved when self-monitoring activities are performed in real-time and proximal to the target behavior [42]. This indicates that mHealth apps could better help users self-monitor health-related behaviors using a notification system that cannot be eliminated by participants to remind them throughout the day and at times in which the behaviors typically occur.

In a qualitative study examining the desired features of weight loss apps among young adults, Tang and colleagues found that participants valued the opportunity to move beyond strictly tracking their eating behaviors and wanted a way to integrate other features, such as behavioral weight loss goals [40]. Other mobile apps have also incorporated a goals feature. Morrison and colleagues instructed participants to set their own goals and track their goal achievement progress using their mobile app and found that participants most frequently accessed informational content using their app (eg, food lists); fewer participants used the goal setting and monitoring features of their mobile app [39]. As some participants are less likely to use app components that require initial set-up (eg, setting goals or notifications), including some type of preset behavioral goals for users to achieve could help promote user engagement and motivation with these mHealth apps.

In their qualitative study documenting the development and prototype testing of an app promoting change in eating and activity behaviors to reduce weight gain among young adults, Hebden and colleagues received feedback requesting positive reinforcement for performing desired behaviors (eg, eating healthy foods or engaging in physical activity) [41]. This is similar to participant suggestions for more opportunities to provide and receive praise in the Social POD study. Including opportunities for users to give and receive praise for performing targeted health behaviors could be another necessary component to help establish and maintain health behavior change within the context of mHealth interventions.

The color scheme of the Social POD app will be updated with brighter colors to better appeal to users as suggested in the Social POD participant focus groups. Tang and colleagues conducted focus groups with young adult participants and also found that the perceived attractiveness of an app was an important consideration for participant satisfaction and maintaining engagement with weight loss apps [40]. Functionality issues similar to those found in the notification system of the Social POD app were found in a study conducted by Morrison and colleagues where a participant reported not receiving notifications during testing of a new mobile app for goal setting and self-monitoring of diet and activity among a small sample of adults [39].

Ensuring that all components of mHealth apps are functioning properly over time is an integral part of the usability testing process and of great importance in remotely delivered behavioral health interventions. Following usability testing, it was imperative to prioritize the changes that were made to the Social POD app prior to the pilot RCT. Correcting functionality issues and developing a newsfeed and incentive systems took priority; an avatar to personalize the app was not incorporated, and, while a nutrient database was added, it was not as extensive as originally hoped.

Limitations

Because the purpose of this study was to test the first iteration of the Social POD app prior to the pilot RCT, it was not adequately powered to detect statistically significant differences in pre-post scores. The time period of the study was limited and at 2 months may have been too short to detect significant differences in participant pre-post body weight. The Social POD app is currently only available for the Android operating system. Despite the fact that the Android phone is the most prevalent cell phone in the United States [43], the fact that other smartphones (eg, iPhone) were excluded may reduce the generalizability of the findings. Direct questioning (versus open-ended questioning) was used to solicit participant opinions regarding the addition of an incentive system, which could have resulted in bias. The sample size of this study was also very small, at just 9 participants and only 1 male, and was therefore not a representative sample of all potential users of the Social POD app. Because the sample size was small, message selection results could be skewed toward the message type that frequent participants preferred. Participants were instructed to use a commercial database to identify the caloric content of food and beverages consumed and use of this database could have contributed to the change in weight observed following this intervention.

Strengths

One strength of this study was the use of a mixed-methods design, which included participant focus groups to obtain usability and functionality feedback and suggestions for improvements to the Social POD app prior to the pilot RCT. Obtaining feedback through the weekly surveys was another

strength, given the potential for some participants to refrain or modify comments during focus groups based on social desirability bias. Another strength of this study was the iterative testing of the Social POD app to uncover and resolve functionality issues prior to the pilot RCT [34]. While the total number of app entries was not statistically significantly correlated with percent weight loss, a correlation coefficient of r=.49 represents a fairly large effect size [44]. Furthermore, mean difference in calories expended during intentional physical activity from pre- to posttest was also not significantly different, but this represented a small effect and could be greater if tested among a larger sample and over a longer time period [44]. While the sample size was small in this usability study, the minimum percentage of problems identified during testing of mobile apps increased from 55% to 85%, respectively, when the sample size increased from 5 participants to 10 [45]. The reach of mHealth interventions such as this has the potential to be even greater than traditional face-to-face interventions, and even small changes in weight have the potential to impact public health outcomes and reduce disease risk [46]. A reduction in body weight as little as 1 kg, as seen in this study, has been associated with a 16% reduction in type 2 diabetes risk [47] demonstrating that this type of mHealth intervention is a scalable way to deliver a weight loss program with beneficial reduction of disease risk. The comparison of objective and subjective reports of app use is another strength of this study.

Conclusion

The Social POD app provides an innovative way to encourage self-monitoring of dietary intake, weight, and physical activity while encouraging frequent users to provide social support to infrequent users. Although more time is needed for development, this mHealth intervention can be disseminated broadly for many years and to many individuals without the need for additional intensive in-person hours. The Social POD app should be tested in a larger clinical trial for a longer length of time to determine if changes in participant weight, calories consumed, and calories expended during physical activity are improved.

Acknowledgments

This project was supported by the South Carolina Clinical and Translational Research Institute at the Medical University of South Carolina, CTSA NIH/NCATS grant UL1TR000062 to Dr Turner-McGrievy, and NIH grants 1R01GM081793 and P20 RR-016461 to Dr Homayoun Valafar. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors would like to thank Lynn Boyd for her contributions to the focus group sessions and transcription of audio recordings.

Conflicts of Interest

None declared.

References

1. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. JAMA 2012 Feb 1;307(5):483-490. [doi: 10.1001/jama.2012.40] [Medline: 22253364]
2. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001. JAMA 2003 Jan 01;289(1). [doi: 10.1001/jama.289.1.76]
3. Must A. The Disease Burden Associated With Overweight and Obesity. JAMA 1999 Oct 27;282(16):1523. [doi: 10.1001/jama.282.16.1523]
4. Ma J, Huang M, Wang L, Ye W, Tong Y, Wang H. Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies. Med Sci Monit 2015;21:283-291 [FREE Full text] [doi: 10.12659/MSM.892035] [Medline: 25612155]
5. Bianchini F, Kaaks R, Vainio H. Overweight, obesity, and cancer risk. The Lancet Oncology 2002 Sep;3(9):565-574. [doi: 10.1016/S1470-2045(02)00849-5]
6. Dombrowski SU, Avenell A, Sniehott FF. Behavioural interventions for obese adults with additional risk factors for morbidity: systematic review of effects on behaviour, weight and disease risk factors. Obes Facts 2010 Dec;3(6):377-396. [doi: 10.1159/000323076] [Medline: 21196792]
7. Sherwood NE, Morton N, Jeffery RW, French SA, Neumark-Sztainer D, Falkner NH. Consumer preferences in format and type of community-based weight control programs. Am J Health Promot 1998;13(1):12-18. [Medline: 10186930]
8. Fox S, Duggan M. Mobile health 2012. Pew Internet Am Life Proj 2012;8.
9. 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]
10. Azar Kristen M J, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, 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]
11. Hales S, Davidson CR, Turner-McGrievy GM. Varying social media message types differentially impacts engagement in a behavioural weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362. [doi: 10.1007/s13142-014-0274-z]

12. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
13. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
14. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
15. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
16. Cardinal BJ, Esters J, Cardinal MK. Evaluation of the revised physical activity readiness questionnaire in older adults. Med Sci Sports Exerc 1996 Apr;28(4):468-472. [Medline: 8778552]
17. Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering. Internet Comput IEEE 2003;7:76-80.
18. Linden G, Conover M, Robertson J. The Netflix prize, computer science outreach, and Japanese mobile phones. Commun. ACM 2009 Oct 01;52(10):8. [doi: 10.1145/1562764.1562769]
19. Pazzani M, Billsus D. Content-Based Recommendation Systems. In: Brusilovsky P, Kobsa A, Nejdi W, editors. The Adaptive Web. Berlin Heidelberg: Springer-Verlag; 2007:325-341.
20. Shani G, Gunawardana A. Evaluating Recommendation Systems. In: Ricci F, Rokach L, Shapira B, Kantor PB, editors. Recommendation Systems Handbook. US: Springer-Verlag; 2011:257-297.
21. Giabbanelli PJ, Crutzen R. Supporting self-management of obesity using a novel game architecture. Health Informatics J 2015 Sep;21(3):223-236. [doi: 10.1177/1460458214521051] [Medline: 24557604]
22. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
23. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Dec 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
24. National Cancer Institute. 2005. Theory at a Glance: A Guide for Health Promotion Practice URL: http://www.cancer.gov/cancertopics/cancerlibrary/theory.pdf [accessed 2016-01-04] [WebCite Cache ID 6eI4IjOHZ]
25. Bandura A. Social foundations of thoughtaction: A social cognitive theory. Englewood Cliffs, NJ. US: Prentice-Hall, Inc; 1986.
26. Anderson-Bill ES, Winett RA, Wojcik JR, Winett SG. Web-based guide to health: relationship of theoretical variables to change in physical activity, nutrition and weight at 16-months. J Med Internet Res 2011 Mar;13(1):e27 [FREE Full text] [doi: 10.2196/jmir.1614] [Medline: 21447470]
27. Turner-McGrievy GM, Campbell MK, Tate DF, Truesdale KP, Bowling JM, Crosby L. Pounds Off Digitally study: a randomized podcasting weight-loss intervention. Am J Prev Med 2009 Oct;37(4):263-269 [FREE Full text] [doi: 10.1016/j.amepre.2009.06.010] [Medline: 19765496]
28. Subar AF, Crafts J, Zimmerman TP, Wilson M, Mittl B, Islam NG, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. J Am Diet Assoc 2010 Jan;110(1):55-64 [FREE Full text] [doi: 10.1016/j.jada.2009.10.007] [Medline: 20102828]
29. Paffenbarger RS, Wing AL, Hyde RT, Jung DL. Physical activity and incidence of hypertension in college alumni. Am J Epidemiol 1983 Mar;117(3):245-257. [Medline: 6829553]
30. Clark MM, Abrams DB, Niaura RS, Eaton CA, Rossi JS. Self-efficacy in weight management. J Consult Clin Psychol 1991 Oct;59(5):739-744. [Medline: 1955608]
31. Benet-Martínez V, John OP. Los Cinco Grandes across cultures and ethnic groups: multitrait multimethod analyses of the Big Five in Spanish and English. J Pers Soc Psychol 1998 Sep;75(3):729-750. [Medline: 9781409]
32. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [Medline: 3432232]
33. Driver SJ. Psychometric properties and analysis of the physical activity Social Influence Scale for adults with traumatic brain injuries. Adapt Phys Activ Q 2007 Apr;24(2):160-177. [Medline: 17916915]
34. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
35. Bastien JM. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010 Apr;79(4):e18-e23. [doi: 10.1016/j.ijmedinf.2008.12.004] [Medline: 19345139]
36. O'Malley G, Dowdall G, Burls A, Perry IJ, Curran N. Exploring the usability of a mobile app for adolescent obesity management. JMIR Mhealth Uhealth 2014 Jun;2(2):e29 [FREE Full text] [doi: 10.2196/mhealth.3262] [Medline: 25098237]
37. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014 May;2:42 [FREE Full text] [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]

38. Santorelli G, Petherick ES, Wright J, Wilson B, Samiei H, Cameron N, et al. Developing prediction equations and a mobile phone application to identify infants at risk of obesity. PLoS One 2013 Aug;8(8):e71183 [FREE Full text] [doi: 10.1371/journal.pone.0071183] [Medline: 23940713]
39. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014 Oct;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
40. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Feb;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
41. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
42. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
43. Smith A. Smartphone Ownership 2013.: Pew Research Center; 2013 Jun 05. URL: http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/ [WebCite Cache ID 6eI4XguIS]
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale: Lawrence Erlbaum; 1988.
45. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [Medline: 14587545]
46. Estabrooks P. Translational Science in the Behavioral Domain: More interventions please..but enough with the efficacy!. 2013 Presented at: Australian Society for Behavioral Medicine; 2013; Newcastle, Australia.
47. Hamman RF, Wing RR, Edelstein SL, Lachin JM, Bray GA, Delahanty L, et al. Effect of weight loss with lifestyle intervention on risk of diabetes. Diabetes Care 2006 Sep;29(9):2102-2107 [FREE Full text] [doi: 10.2337/dc06-0560] [Medline: 16936160]

Abbreviations

ASA24: Automated Self-Administered 24-hour Dietary Recall
BMI: body mass index
PAR-Q: Physical Activity Readiness Questionnaire
RCT: randomized clinical trial
Social POD: Social Pounds Off Digitally
SCT: social cognitive theory

Edited by G Eysenbach; submitted 09.04.15; peer-reviewed by L Morrison, L Yingling, G O'Malley; comments to author 19.08.15; revised version received 14.09.15; accepted 02.11.15; published 12.02.16.

Please cite as:
Hales S, Turner-McGrievy G, Fahim A, Freix A, Wilcox S, Davis RE, Huhns M, Valafar H
A Mixed-Methods Approach to the Development, Refinement, and Pilot Testing of Social Networks for Improving Healthy Behaviors
JMIR Human Factors 2016;3(1):e8
URL: http://humanfactors.jmir.org/2016/1/e8/
doi:10.2196/humanfactors.4512
PMID:27026535

Original Paper

Implementation of a Computerized Screening Inventory: Improved Usability Through Iterative Testing and Modification

Edwin D Boudreaux[1], PhD; Andrew Christopher Fischer[2], MD; Brianna Lyn Haskins[3], MS; Zubair Saeed Zafar[4], MD; Guanling Chen[5], PhD; Sneha A Chinai[3], MD

[1]Departments of Emergency Medicine, Psychiatry, and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[2]Department of Emergency Medicine, Kern Medical Center, Bakersfield, CA, United States

[3]Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[4]Department of Internal Medicine, Southern Illinois University School of Medicine, Springfield, IL, United States

[5]Department of Computer Science, University of Massachusetts, Lowell, MA, United States

Corresponding Author:
Edwin D Boudreaux, PhD
Departments of Emergency Medicine, Psychiatry, and Quantitative Health Sciences
University of Massachusetts Medical School
55 Lake Avenue North
Worcester, MA, 01655
United States
Phone: 1 508 334 3817
Fax: 1 508 421 1490
Email: edwin.boudreaux@umassmed.edu

Abstract

Background: The administration of health screeners in a hospital setting has traditionally required (1) clinicians to ask questions and log answers, which can be time consuming and susceptible to error, or (2) patients to complete paper-and-pencil surveys, which require third-party entry of information into the electronic health record and can be vulnerable to error and misinterpretation. A highly promising method that avoids these limitations and bypasses third-party interpretation is direct entry via a computerized inventory.

Objective: To (1) computerize medical and behavioral health screening for use in general medical settings, (2) optimize patient acceptability and feasibility through iterative usability testing and modification cycles, and (3) examine how age relates to usability.

Methods: A computerized version of 15 screeners, including behavioral health screeners recommended by a National Institutes of Health Office of Behavioral and Social Sciences Research collaborative workgroup, was subjected to systematic usability testing and iterative modification. Consecutive adult, English-speaking patients seeking treatment in an urban emergency department were enrolled. Acceptability was defined as (1) the percentage of eligible patients who agreed to take the assessment (initiation rate) and (2) average satisfaction with the assessment (satisfaction rate). Feasibility was defined as the percentage of the screening items completed by those who initiated the assessment (completion rate). Chi-square tests, analyses of variance, and Pearson correlations were used to detect whether improvements in initiation, satisfaction, and completion rates were seen over time and to examine the relation between age and outcomes.

Results: Of 2157 eligible patients approached, 1280 agreed to complete the screening (initiation rate=59.34%). Statistically significant increases were observed over time in satisfaction ($F_{3,1061}$=3.35, P=.019) and completion rates ($F_{3,1276}$=25.44, P<.001). Younger age was associated with greater initiation (initiated, mean [SD], 46.6 [18.7] years; declined: 53.0 [19.5] years, $t_{2,155}$=−7.6, P<.001), higher completion (r=−.20, P<.001), and stronger satisfaction (r=−.23, P<.001).

Conclusions: In a rapid-paced emergency department with a heterogeneous patient population, 59.34% (1280/2157) of all eligible patients initiated the computerized screener with a completion rate reaching over 90%. Usability testing revealed several critical principles for maximizing usability of the computerized medical and behavioral health screeners used in this study. Further work is needed to identify usability issues pertaining to other screeners, racially and ethnically diverse patient groups, and different health care settings.

(JMIR Human Factors 2016;3(1):e10) doi:10.2196/humanfactors.4896

KEYWORDS

behavioral medicine; computers; electronic health records; public health; screening; telemedicine

Introduction

Electronic health records (EHRs) have become ubiquitous in health care settings, but their full capacity to markedly improve public health has yet to be realized. To catalyze the transition from "public health potential" to "public health improvement," a landmark collaboration supported by the National Institutes of Health (NIH) Office of Behavioral and Social Sciences Research (OBSSR) has derived a list of recommended core psychosocial screeners to be incorporated into *all* EHRs [1]. The multiphase consensus building process used by the OBSSR collaborative workgroup included diverse stakeholders from health care systems, scientists, policy makers, governmental organizations, health insurers, clinicians, and consumer groups. To facilitate downstream implementation, the panel sought to ensure that the recommended screeners (see Multimedia Appendix 1) were not only rooted in strong science but were also actionable, user friendly, clinically relevant, and cost effective. Adoption of these core measures into EHRs could improve individual-level patient care, identify drivers of hospital readmissions, and facilitate public health research by supporting more efficient, accurate harmonization of data across different EHRs.

Even if an EHR company integrates the core behavioral health screeners into its templates, clinicians still must ask the questions and enter the answers, a task that can be time consuming and susceptible to error, especially when multiple screeners spanning a variety of domains must be administered. Even if paper-and-pencil administration is used by the patient to complete the screeners, the clinician or a designee still must enter the item responses or scale scores into the EHR manually. A highly promising method that avoids these limitations and bypasses third-party interpretation leading to potential misinterpretation is direct entry via computerized assessments. In many situations, electronic collection of screeners is superior to verbal interview because it guarantees standardized administration of the questions and scoring, promotes forthcoming responses by reducing social desirability bias [2-4], and requires less clinician time. Because data from computerized assessments has the potential to be imported directly into the EHR, it can reduce transcription and scoring errors and time associated with manual entry of paper-based item responses or scale scores. A truly integrated system that pairs computerized self-assessments of the OBSSR core screeners such that the data output matches up precisely with the same template fields in the EHR would be a strong innovation. Advances in tablet computing make such integration in clinical settings even more practical, because of the ease of administration, low cost, and growing familiarity with the medium in the general population.

Behavioral health screeners are not the only important information to obtain. From the medical provider's perspective, computerizing the OBSSR behavioral health screeners alone has limited utility. When managing a heterogeneous group of patients in a general medical setting, like a primary care clinic or an emergency department, screening for medical symptoms, conditions, and diagnoses (eg, pain, chronic illnesses, and surgical history) is equally important. Consequently, for a computerized screening inventory to have optimal utility in general medical settings, it should assist with screening for both medical and behavioral health domains and be integrated with the EHR.

The overall effectiveness of an integrated EHR-computerized screening system assessing both medical and behavioral health statuses will depend on numerous factors. One of the most important is the system's acceptability and feasibility among patients receiving care. It is essential to design the items and user interface to maximize patient usability. In this context, usability relates to how easy the computerized assessment is to complete [5]. Typically, during usability testing, representative participants are asked to complete the assessment in a manner similar to the intended deployment while trained research staff observe and debrief participants. In addition, data collected by the computerized assessment can be used to evaluate usability, such as examining patterns in missing data to determine challenging items. Usability testing should identify problems that impede successful completion and collect qualitative and quantitative data that help the team to understand the root causes of these impediments. Furthermore, the best usability studies not only identify these impediments but also systematically attempt to remediate them by modifying the items, interface, or administration procedures by evaluating the resulting impact on usability in an iterative fashion.

Although much has been written on designing usable websites from a commercial perspective [6], the literature on usability of computerized screeners designed for use in general medical settings is quite limited. For example, Hess and colleagues [7] published data on more than 11,000 administrations of a tablet-based patient self-assessment in a primary care practice and showed that 84% reported no difficulty in completing the assessment. However, they did not report the proportion of the total population that agreed to complete the computerized assessment, or initiation rate, nor did they obtain systematic information on impediments to completion that may have been used to further improve acceptability and feasibility. While 84% may seem like a strong performance, in busy clinical settings it may be unacceptable, because it suggests that 16% may either be dissatisfied or report problems to clinical staff, who do not have the time or the training to address such issues.

In addition to a general lack of rigorous research on usability of computerized screeners, the association between age and usability remains poorly understood. Some studies have shown age to be inversely associated with usability of computerized assessments [7,8], whereas others have not [9]. The relation between age and usability is important to understand because it could introduce systematic bias into both the clinical monitoring of health behaviors among patients and the public

health research that utilizes data resulting from these assessments. Research is needed to better understand how age relates to computerized screener acceptance and feasibility.

The aims of the current study were to (1) computerize a core set of medical and behavioral health screeners, (2) optimize patient acceptability and feasibility through iterative usability testing and modification cycles with a sample of heterogeneous medical patients, and (3) examine how age is associated with patient acceptance and feasibility.

Methods

Study Setting

The study was set in a large, urban, academic emergency department, which is a good setting for usability testing of a computerized screening inventory for several reasons. First, because of the nature of emergency care, providers know little about the patients when they arrive. Screening for pain and other past medical history is important.

Second, broad mandates to incorporate behavioral health screening efforts into emergency care exist [10]. This is true for the following reasons: (1) many patients do not have access to primary care, so if behavioral health is not addressed in the emergency department it is often not addressed at all, and (2) many presentations are directly related to health behaviors, such as an automobile crash resulting from driving while intoxicated. Consequently, the emergency department is an important setting in its own right for preventive health efforts. The OBSSR screeners are a particularly good fit for the emergency department because they are very brief, with only 1 or 2 items per screener.

Third, patient volume is brisk and large samples needed for iterative cycles can be generated quickly. The nature of emergency department care allows for patients to have downtime to complete the assessment while they wait for clinician evaluation, test results, consultants, or inpatient beds.

Participants

From January to December 2013, data collection shifts represented 7 days of the week and ranged between 9 am and 10 pm. During each research shift, every patient who presented for care in the emergency department was logged and considered for participation regardless of presenting complaint to maximize sample representativeness. Patients were excluded if they were younger than 18 years of age, non-English speaking, incarcerated, or medically, cognitively, or emotionally unable to be interviewed or to respond to a computerized assessment (eg, intubation, persistent vomiting, severe pain, altered mental status). Of the 5000 patients logged, 2592 (51.84%) were interviewed by research assistants (RAs); the others were not interviewed due to exclusion criteria (see the "Study Procedure" section), patient unavailability, or research staff unavailability. Of those interviewed, 2157 (83.22%) were deemed eligible; of these, 1280 (59.34%) agreed to take the assessment. The mean (SD) age of the consenting sample was 46 (17) years, and 555 (43.35%) were women, 1021 (79.77%) were white, and 60 (4.69%) were Hispanic.

Study Procedure

A multidisciplinary team composed of a health psychologist, physicians, nurses, a nurse practitioner, and computer scientists helped build the initial specifications for the computerized screening inventory. The inventory (Vecna Technologies, Inc, Cambridge, MA, USA) is Web-based, hosted on a server compliant with the Health Information Portability and Accountability Act, and designed to be presented on a tablet. The project team created medical screening items that were deemed most important to the emergency department setting. These included pain (intensity and location), other medical symptoms associated with pain, and past medical, psychiatric, and surgical history of the patients (see Multimedia Appendix 1). The OBSSR behavioral health screeners were included, as well as follow-up items in response to positive screens, where appropriate, such as the type of illicit drugs used if the individual screened positive for use. Longer follow-up screeners, such as the Alcohol Use Disorders Identification Test (AUDIT) [11], are recommended by OBSSR for positive initial screens. However, these longer screeners were not included to preserve the feasibility of the administration. The medical items were presented first because the team thought this would promote perceived relevance of completing a computerized screening because most patients present to the emergency department for medical not behavioral complaints.

For the original deployment, the computerized screening inventory was designed to mimic paper intake forms routinely used in medical settings. Multiple items appeared on the screen, and patients indicated their answers by touching the response options and scrolling down to access the rest of the items on the page. Upon completing their current page, patients tapped the "Next" button, and were presented with the next multi-item page. The project team believed that this would be a highly efficient administration format that aligned with a paper-based process with which patients were already familiar with, thereby improving acceptability. The format of the items' response options was initially allowed to vary based on the particular item. For example, the response to the tobacco use question was binary (Yes/No), whereas illicit drug use was numeric (the number of days in the past 12 months drugs were used). This aligned with the published OBSSR screeners. Patients could skip items at will, a feature the team believed would respect patient's autonomy by allowing the individual to skip questions he/she did not want to answer. The assessment administration ended automatically after the final answer was entered.

All items and responses used the same font style and size to maintain consistency. All items and instructions were framed in the second person. Because it is difficult to make adjustments simultaneously across numerous languages, only an English version was tested. The project team intends to translate and test the final version with other groups in future studies. The minimum number of items presented was 37 (see Multimedia Appendix 1); the maximum, counting all branched items, was 41 items. Because some screeners required more than 1 item, there are more items than screeners. The computerized screening inventory was extensively tested by the project team, debugged by the Vecna engineers, and piloted with an initial sample of 20 patients prior to full patient testing. Modifications to the base

system resulting from usability testing are described in the "Results" section.

Trained RAs first determined if an individual should be excluded through a combination of medical chart review and brief discussion with the treating clinicians. Those clearly satisfying exclusion criteria, such as patients who were being resuscitated, documented as non-English speakers, incarcerated, or physically incapable of completing the electronic assessment were excluded. The rest were approached at the bedside. Approach and consent were concise to make the experience as naturalistic as possible. Following a brief introduction, the RA asked the patient if he/she was willing to participate in a study that involved answering health-related questions on a tablet. The RA assured the patient that experience with computers or tablets was not necessary, their medical care would not be interrupted or delayed by participation, and they could stop at any time. Interested patients provided verbal consent.

For those who consented, the RA opened the computerized screening inventory on the tablet, provided basic instructions, handed the tablet to the patient, and remained present for the first few demographics items (eg, name, age) to ensure the patient understood how to proceed. After the first few items, the RA left the patient's bedside to provide privacy but remained nearby in case the patient required assistance or was interrupted for medical care. After the patient completed the computerized screening inventory, he/she reviewed a summary of his/her answers for accuracy, and errors were corrected. The RA concluded by performing a semistructured debriefing interview that assessed perceived barriers, challenges, and suggested improvements to the system. The tablet was housed in a protective case and sanitized after every patient administration for infection control.

The RA documented all questions and problems observed throughout the enrollment process, including results from the debriefing interview, on a patient experience log (described in the "Measures" section). This log was summarized by research staff on a weekly basis and reviewed by the principal investigator and other members of the research team. Recommended system changes were identified, prioritized based on their likely impact on usability, and shared with the vendor. Each update to the computerized screening inventory was debugged and tested by a quality assurance team prior to release. Testing, problem identification, and further modifications continued systematically throughout the study period. In addition to changes to the inventory, problems related to the RA's introduction and administration procedures were identified and solutions implemented. Although small iterative refinements in the software, item content, and administration procedures were made throughout the study, major clusters of changes occurred at 3 time points, which divided the 12 months into 4 phases (see the "Results" section).

The study was approved by the UMass Institutional Review Board, in accordance with all applicable regulations, and informed consent was obtained after the nature and possible consequences of the study were explained.

Measures

Demographics

Age, sex, race (white vs. nonwhite), and ethnicity (Hispanic, non-Hispanic) were documented for all patients approached during the research shifts.

Computerized Screening Inventory

The inventory initially consisted of 41 possible items. The medical items were created through team consensus because a standardized medical screening form suitable for the emergency department could not be identified in the literature. Items and response options of the OBSSR screeners followed Estabrooks and colleagues [1], with 2 exceptions. The single-item alcohol screener, "How many times in the past year have you had "X" or more drinks in a day (where "X" is 5 for men, 4 for women)?" was replaced by the 3-item AUDIT-C [12]. The AUDIT-C has been validated in the emergency department setting, whereas the single-item screener has not yet been. The AUDIT-C has an item to assess binge drinking that is very similar to the single-item OBSSR screener, so the computerized screening inventory covered the OBSSR-recommended screening plus 2 items assessing average weekly consumption.

The second deviation pertained to the stress thermometer. Estabrooks and colleagues [1] referred to a "stress" thermometer but used the word "distress" in the item. The "distress" thermometer has never been validated in an emergency department setting and the study team felt that patients would better understand the word "stress," so it was used instead.

Usability

All RAs made objective observations of the entire administration of the computerized screening inventory, from the initial opening of the inventory to the debriefing interviews. All observations were documented on the patient experience log. This included those observed directly by the RA and those reported by the patient during debriefing. Detailed descriptions of problems were prepared, including representative case studies for team review. Overall completion rates and item skip patterns were summarized intermittently to complement the patient experience log summaries.

Acceptability

Patient acceptability was measured by 2 indicators. First, the "initiation rate" was defined as the number of patients who agreed to take the survey divided by the number of patients who were eligible. Second, the "satisfaction rate" was an average of 3 items administered at the end of the inventory: (1) assessment length ("much too long," "a little too long," "about right," "a little too short," and "much too short"), (2) ease of understanding the items ("very difficult," "somewhat difficult," "neither difficult nor easy," "somewhat easy," and "very easy"), and (3) ease of using the tablet ("very difficult," "somewhat difficult," "neither difficult nor easy," "somewhat easy," and "very easy"). The ratings were averaged to create an overall satisfaction score, with higher scores reflecting stronger satisfaction (range 0-4).

Feasibility

The operational definition of feasibility was the percentage of the survey that an individual completed, or the completion rate.

Completion of a screener was counted only if enough information was provided to accurately determine if the patient was positive or negative for the condition. For multi-item screeners, this meant all items had to be answered. In all, 3 completion rates were derived. The overall completion was defined as, among those patients who agreed to participate, the number of screeners completed divided by 15 (the total possible screeners). Medical completion was defined as, among those patients who agreed to participate, the number of medical screeners completed divided by 9 (total number of medical screeners in the inventory). Behavioral health completion was defined as, among those patients who agreed to participate, the number of behavioral health screeners completed divided by 6. Only the 6 behavioral health screeners administered throughout the entire study were used to maintain a consistent denominator across the study. The completion rates ranged from 0% (for a person who agreed to take the survey but did not complete a single screener) to 100% (for a person who completed all of the screeners).

Data Analytic Plan

Changes in initiation rate (Yes/No, categorical), average satisfaction (continuous), and completion rates (continuous) over time were examined using chi-square tests and one-way analysis of variance (ANOVA), with phase (defined by major upgrades/changes) as the independent variable. Associations between age and outcomes were examined using ANOVAs, independent samples *t* tests, chi-square tests, and Pearson correlations. All data were analyzed using Statistical Package for the Social Science 22 (IBM, Armonk, NY, USA).

Results

Usability Testing and Modifications

Table 1 summarizes the major usability problems noted and the resulting changes in the system and administration procedures that were made. Major modifications to the system or administration procedures occurred at 3 points, which split the study into 4 phases. The first major change, which delineated Phase 1 from Phase 2, updated the user interface to use larger font, bolded key phrases, improved contrast between background and items, provided better space separation between items and responses, and presented fewer questions on the screen to eliminate the need for scrolling. Greater clarity on how to navigate the system was added to the RA instructions and the screens, such as how to access the numeric keypad when an integer was needed for a response. In addition, all primary items, or items that were presented to all individuals and which were not branched based on the response to an earlier item, became required rather than allowing "skipping at will" to improve confidence that items with missing data were intentionally skipped. The second major upgrade, which delineated Phase 2 from Phase 3, included presenting a single item per page (rather than multi-item pages), adding "Do not understand" and "Prefer not to answer" to all required items, optimizing the look and feel for tablet presentation, changing all integer response fields to multiple choice "buttons," and adding the capability of easily editing the items from a final "Confirmation" screen. The total length was shortened by removing 15 items, leaving a total of 29 items. This included removing 3 of the OBSSR screeners that were judged to be less important for the emergency care setting (7 items assessing diet, exercise, sleep) and 8 items assessing demographics. The final major change, which delineated Phase 3 from Phase 4, included adding instructions to help prevent "double tapping" while the Web page was being refreshed between items, which was resulting in some items being inadvertently skipped.

Table 1. Usability impediments and solutions applied.

Problem description	Solution applied
1. Technical	
Disrupted Internet connectivity resulted in "frozen assessments" and lost data	Wi-Fi system upgrades (coincidental to the study).
	Tablets were paired with the Clinical Wi-Fi, rather than the Guest Wi-Fi, to improve reliability.
	Staff members were trained to ensure Wi-Fi connection at the beginning of each shift.
	Staff were trained to avoid opening the computerized screening inventory until it was needed to avoid browser time-outs associated with long dormant times.
2. Survey content/item structure	
Survey length prompted discontinued and interrupted assessments, as well as some patient dissatisfaction	The team chose to remove items that were deemed less relevant for the setting and demographics that would likely be already collected in the electronic health record, thereby shortening the total length (from 41 to 26 primary items).
Integer responses requiring numeric keypad entry were problematic because of skill/knowledge required for accessing the touch screen numeric keypad	All response options were changed to categorical "buttons" (ie, free-text integer responses were eliminated for all items).
Some patients had trouble understanding or did not want to answer some items	We added 2 response options to every primary item: "Do not understand" and "Prefer not to answer."
3. User interface/layout	
Skipped items/missing data resulting from multi-item "form" layout (eg, it was difficult to clearly differentiate between items because they were too close together and were skipped, scrolling down to get to the next items led to the patient inadvertently skipping items because they scrolled past them and did not realize it)	Changed from a multi-item "form" based administration to a single item per page.
	Font maximized for single-item presentation.
	No scrolling required.
	Spacing and color contrast were adjusted to maximize differentiation between the item and response options from the background, the item stem from the response options, and the response options from each other.
	Open-response format where patients could skip questions "at will" changed to requiring an answer prior to proceeding to the next question.
Users sometimes responded to questions but did not realize that they had "tapped" the wrong response until they reviewed the summary of their responses during the debriefing	A final screen was added that allowed the patient to easily review their answers to all of the items and "Confirm" the answers were correct, or easily go back to an item to edit if needed.
4. Administration process and instructions	
Lack of familiarity with touch screen interface created difficulty while navigating and skipped items	
	Opening instructions were modified to be more specific to training patients to understand the basics of responding on a touch screen, including how to scroll and the importance of waiting after tapping a response to avoid double-tapping.
	The option of using a stylus was provided.
	The option of propping the tablet on a tray table was added to help patients who were having trouble holding the tablet (eg, elderly, frail patients).
Patients could not complete the survey themselves and requested assistance	Family members or friends accompanying the patient could complete the assessment on their behalf (proxy assessment).
Assessments were interrupted frequently by medical testing, procedures, visitors, etc	A time out and "pause" feature that closes the browser while saving data and allowing resumption from the item last completed was implemented.

Of the 1280 administrations, 61 (4.77%) had a significant technical problem, primarily Wi-Fi interruption; 238 (18.59%) had a usability issue related to the interface, such as problems scrolling or accessing the numeric keypad, although the vast majority of these issues did not prevent the individual from completing the assessment; 411 (32.11%) were interrupted by

medical testing, procedures, visitors, or other reasons; and 162 (12.66%) had a family or friend (proxy) complete the assessment on behalf of the patient.

Acceptability: Initiation

Of the 2592 emergency department patients approached by research staff, 2157 were deemed eligible. Among those eligible, 877 (40.66%) declined and 1280 agreed to participate, for an overall initiation rate of 59.34%. The initiation rate did not differ statistically over the 4 phases, χ_3^2 (N=2157) = 8.69, P>.05. Those who initiated mean [SD] the survey (46.6 [18.7] years) were younger, on average, than those who declined (53.0 [19.5] years), $t_{2,155}$=−7.6, P<.001.

Acceptability: Satisfaction

A one-way ANOVA revealed statistically different average satisfaction rates between phases, $F_{3,1061}$=3.35, P=.019, with Tukey post hoc tests revealing that satisfaction (mean [SD]) during Phase 3 (3.10 [0.47]) was significantly higher than that during Phase 2 (2.99 [0.57]). Younger age was associated with stronger satisfaction (r=−.23, P<.001).

Feasibility: Completion

Figure 1 depicts the 3 completion rates (overall, medical, and behavioral) among those who initiated the survey over the 12 months of the study. A one-way ANOVA revealed statistically different average overall completion rates (ie, average percentage of the screeners that were completed) between phases ($F_{3,1276}$=25.44, P<.001). Tukey post hoc tests revealed that Phase 1 (mean [SD] 75% [38%]) and Phase 2 (79% [35%]) were significantly lower than Phase 3 (87% [30%]), which was, in turn, significantly lower than Phase 4 (94% [19%]). Medical screener completion followed a similar pattern, $F_{3,1276}$=23.84, P<.001, as did behavioral screener completion, $F_{3,1276}$=23.57, P<.001. Age was inversely correlated with overall completion (r=−.20, P<.001), medical screener completion (r=−.18, P<.001), and behavioral screener completion (r=−.20, P<.001). The results for each of the screeners, including skip rates, are presented in Multimedia Appendix 2. Multimedia Appendix 2 differs from Multimedia Appendix 1 in that the latter presents all of the individual items administered at the beginning of the study and links them to the screeners with which they are associated, whereas Multimedia Appendix 2 summarizes results pertaining to only 15 screeners that were administered across the entire study.

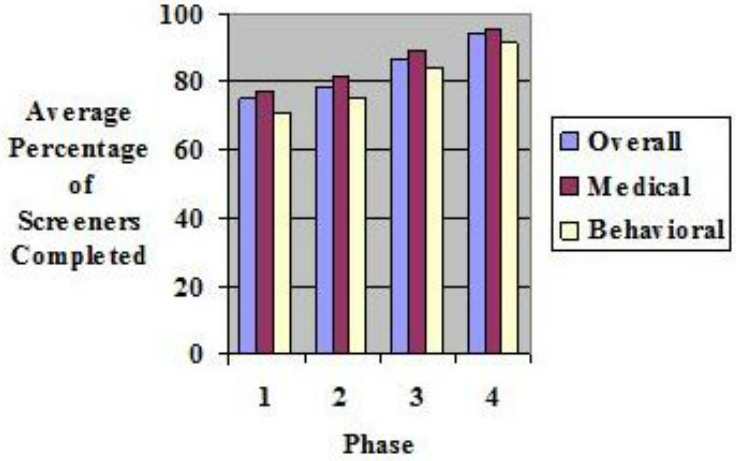
Overall, 15 screeners were administered. Only the 6 behavioral health screeners that were included throughout the entire study are used to facilitate cross-phase comparison (tobacco, risky alcohol consumption, illicit drug use, depression, anxiety, stress). The phases are defined as follows:

Between Phase 1 and 2=Enlarged text size, bolded key phrases in items, better color contrast between items and background, increased space separation between items and responses, presented fewer questions on the screen, eliminated need for scrolling, greater clarity navigating the system was added to the RA instructions and the computerized screening inventory screens, such as instructions on how to access numeric keypad, all primary questions became required.

Between Phase 2 and 3=Presented a single item per page, added “Do not understand” and “Prefer not to answer” options to all required items, optimized the user interface for tablet presentation, changed all integer response fields to multiple choice buttons obviating need for numeric keypad, added the capability of easily editing the items from the final confirmation screen, shortened by removing 15 items (reduced to a total of 26 primary items).

Between Phase 3 and 4=Added instructions to help prevent “double tapping” while the Web page was being refreshed, which was resulting in some items being inadvertently skipped.

Figure 1. Screener completion rates across the study.



Discussion

Preliminary Findings

This is the first study of a computerized screening inventory that blends medical screening items with the NIH OBSSR collaborative’s recommended behavioral health screeners. It is also one of the largest systematic usability studies ever to be conducted on such a system in a general medical setting. The need for systematically testing usability, implementing changes, and testing the effects of these changes was confirmed by the transformative changes that occurred over the course of the study as a result of observations and feedback from patients.

These changes were associated with markedly improved completion rates. By the final phase of the study, average overall completion for those who agreed to take the assessment had risen from 75% to exceeding 90%. This result should be interpreted within the demanding context of the setting. Emergency department patients are often uncomfortable, acutely ill, highly heterogeneous, and interrupted frequently due to medical testing and treatment. All of these factors work against survey completion. If rates over 90% can be achieved in the highly demanding emergency department setting, similar or better results can likely be obtained in other more hospitable health care settings, such as primary care. In addition to improved completion rates, overall satisfaction with the computerized screening inventory improved over time, with the biggest improvement observed between Phase 2 and Phase 3 when the survey was shortened and simplified to a single item per page. By contrast, while completion and satisfaction improved, initial agreement to complete the computerized survey remained roughly stable throughout the study at 59.34% (n=1,280/2,157) of all patients who were eligible. The stability of this indicator is not surprising. Other than reassurances that were provided to the patient from the very beginning of the study, such as no computer experience was needed and the assessment would not delay their medical care, strategies to encourage patients to *begin* a computerized assessment are limited. In addition, the acceptance rate may have been suppressed by the fact that this was introduced as a research study, not as part of care. They may have viewed the experience as something that is not essential to their care. If implemented as part of the standardized care, acceptance rates may actually increase.

The lessons learned in this study likely have implications for other applications of computerized screening and assessment, not just those included in this computerized screening inventory, because many of the barriers were nonspecific to the particular items. The most important problems and the associated solutions are reviewed in the following sections.

Technical Problems

The primary technical barrier centered around the use of Wi-Fi on portable tablets. Tablets are quite popular and are gaining traction in health care settings [13]. Their low cost, portability, and familiarity promote their usability. However, maintaining Internet connectivity when moving from one room to another can be challenging, especially when Wi-Fi capabilities are stressed during peak demand hours and when signals experience interference due to structural barriers. Lost connectivity was the root cause of many of the original “frozen assessments” and lost data. It led to not only entire assessments being lost but also loss of individual questions within an assessment as well. Lost connectivity was made worse by designing the system to avoid caching (temporary storage) of data on the tablets because of data security policies that discourage caching. Interrupted connectivity became less of a problem when the health care system upgraded its Wi-Fi. In addition, research staff training was enhanced. Multiple Wi-Fi networks were available, some with stronger signals than others. RAs were taught how to identify when connectivity to the preferred Wi-Fi had been lost and how to reconnect. Finally, they were trained to log out of the computerized screening inventory completely at the end of the day and to avoid keeping the program open while not in use to avoid browser time outs. Technical solutions that rely on caching, or temporarily storing, data on the hardware and uploading when the connection is restored should also be considered.

Survey Content

A key challenge that computerized screening can help with is the infeasibility of screening for the plethora of recommended screening domains that exist. While computerization represents a potential solution to this problem, a multidimensional computerized assessment still necessitates more items, which leads to longer administration times. Although many patients tolerated the original 41-item survey quite well, a significant portion were interrupted by medical testing, which made them less likely to complete the assessment. In addition, some patients initially complained it was too long. Even a small percentage of dissatisfied patients can dissuade clinicians from adopting a system like the computerized screening inventory. As a result, the total length was shortened by 37% (n=15/41). There is no optimal length for the number of items a computerized assessment should contain, because it is dependent on a host of factors, some of which relate to the assessment objectives, setting, and population. Tolerance for longer assessments may be better in environments with patients who are not as ill as emergency department patients and care processes that are not characterized by frequent, intermittent medical testing and procedures. Careful testing of the acceptability limits and prioritization of the domains assessed are essential for establishing the optimal length in any setting.

Another important finding pertained to item response formats. The recommended wording and response format for several OBSSR screeners necessitated responding with a free-text integer, such as reporting the number of days one used drugs in the past 12 months. However, entering numeric responses challenged some tablet-naïve patients. It required knowing how to access the numeric keypad, which is not immediately obvious and requires knowledge of the correct button to press to activate it. Consequently, the response format was changed for all items to a categorical, button response modality. For example, the illicit drug item was changed from assessing the number of days in the past 12 months that drugs were used to assessing whether the individual had used any drugs in the past 12 months, Yes/No. This provided for a consistent, categorical response set throughout the assessment, rather than switching back and forth from categorical responses to numeric responses, and avoided any need to access the numeric keyboard, which made completing the assessment easier. Notably, at least four of the OBSSR screeners that use numeric response options (see Multimedia Appendix 1) may need to be adapted when computerized. The impact of this modification on reliability and validity is unknown and may need to be tested prospectively.

User Interface

Some usability issues were rooted in user interface design choices. Initially, the team sought to design a highly efficient interface that presented multiple items similar to a paper-based form, thereby presenting the information in a format familiar

to patients and reducing the number of page turns the individual had to complete. However, this multi-item format was text dense, required smaller font, and the items and responses were spaced too closely together. As a result, patients more easily passed over items, especially while scrolling, or mistakenly selected options near to the intended target. Even when the format changed to remove the need for scrolling yet maintaining the multi-item format by presenting fewer items on the page, some patients, especially those with vision problems, still had difficulty reading the text. Consequently, the interface was ultimately changed to present a single item on a page, which allowed marked improvements in font size and spacing. While this increased the number of page turns needed, it helped to prevent inadvertent skips and promoted accurate response selection.

Because health screenings can assess potentially sensitive topics, like alcohol and drug use, it is important to respect patient autonomy to refuse to answer. Initially, patients could freely skip items if they did not want to respond. However, it was impossible to determine if the missing data were deliberate (the patient did not want to answer the question) or inadvertent (the patient did not see the item). This was addressed by adding 2 response options, “Do not understand” and “Prefer not to answer” to all primary items. This allowed patients to decline to answer an item while still requiring a response to each item, thereby removing the ambiguity around missing data, and helped to flag items that were either poorly worded or potentially sensitive.

The overall item look-and-feel on the page was very important. The design principles that emerged can be summarized as follows: maximize the font size to improve readability, maintain strong differentiation between the item stem and the response options, maintain good spacing between the response options, and allow for the entire response text to be “active” such that touching any part of the response is sufficient to enter a response. These user interface design features are particularly important for visually impaired patients or patients who have fine motor skills impairment that might impede their ability to accurately touch their intended response option. Radio buttons alone, a common response entry method used in computerized surveys, were woefully inadequate.

One additional design feature that is important to highlight is the confirmation process at the end of the assessment. Simply concluding the assessment after the individual completes the final survey item can result in erroneous responses going unnoticed and, ultimately, entered into the individual’s permanent medical record. Incorporating a final screen that allows the patient to review his or her responses and easily edit incorrect values is an important validation step.

Administration and Instructions

Many screeners, like the OBSSR screeners, are designed for self-administration. However, implementation of computerized screening inventories will have to account for proxy completion, because many users, especially the very ill, elderly, visually impaired, or tablet naïve, preferred to have an accompanying family member or friend complete the assessment for them. To the extent that behavioral health screeners have not been studied for proxy administration, this introduces an unknown source of potential bias in the results. Nevertheless, it clearly improves the usability of the system. Many of the individuals approached would likely not have accepted the offer or completed the assessment if their family or friends had not been allowed to help.

Another practical administration issue that has important design implications pertains to interrupted assessments. In this study, interruptions were frequent, occurring in 32% (n=411/1280) of patients. This was directly related to the nature of care in the emergency department setting, which is characterized by numerous interactions with various health care professionals, medical testing, and treatment procedures. However, interruptions can occur in any medical setting. As a result, computerized assessments require the following features to accommodate interruptions: First, the patient (or proxy) should be able to pause the assessment by clicking a pause button. Second, the system should have a time out feature that saves data and closes the assessment after a period of inactivity. Third, the individual should have the ability to easily resume the assessment from where he or she left off at any point during care.

Age

Age was inversely associated with initial acceptance, completion, and satisfaction. This creates a cumulative effect of completers being over-represented by younger patients. Hess and colleagues [7] found similar results in primary care. The practical impact of this trend is that alternative methods of gathering the data captured by a computerized assessment will be more commonly used with elderly patients. Allowing proxy completion may partially help adjust for this problem.

Limitations

The study was set in an emergency department. While this setting is important in its own right for health behavior screening, and there were practical advantages to performing usability testing in this setting, it may have characteristics that can reduce acceptance and feasibility. This includes high patient acuity and frequent interruptions. Thus, further testing of the computerized screening inventory or similar systems in other medical settings is important. The sample may under-represent minority patients. Additional study on the use of computerized screening batteries with nonwhite, non-English speaking patients is needed.

Of the 2157 patients eligible, 877 (40.66%) declined to initiate the assessment. Importantly, the demographics of those who accepted were very similar to those of the general population, with the exception of age (those who initiated were younger). Because of the relatively large sample, the staffing of RA shifts across all 7 days of the week covering morning, afternoon, and evening hours, and the protocols requiring consecutive consideration of all adult emergency department patients, the sample is highly representative of the population from which it was drawn.

The system did not present the screening questions using audio, which might have led to improved completion by those with poor literacy or eyesight. Audio is difficult in the emergency

department because of competing noise and difficulty providing headphones for patients in an efficient, infection-controlled manner, which led the team to reject this option for this project.

Some of the wording and response options of the OBSSR screeners were modified from the original publication. This limitation is partially mitigated by the preliminary nature of the original OBSSR recommendations, which were intended to prompt further research such as in this study. In addition, most of the implications for developing a usable behavioral health screening system derived from this study are independent of the specific wording of the items.

For a system like the computerized screening inventory to work clinically, both patients and clinicians will need to embrace it. This study did not test clinician acceptability. The research team wanted to focus on patient usability as the first step and intends to explore clinician acceptability and feasibility next. This is important because there are significant challenges, including EHR integration, data visualization and actionable presentation of results, clinician training, workflow modification, and hardware availability and security.

Conclusion

This study focused on a single administration of a multi-item, computerized screening inventory that included items developed for emergency medicine by the study team and behavioral screening items collaboratively developed by the NIH OBSSR for wide use. It incorporated sequential phases of evaluation and refinement that allowed statistical comparison to determine whether changes in content, design, functionality, and training actually resulted in improved usability. Study staff members were trained and dedicated to the study and thus, by design, any loss of interest or commitment by clinicians in administering the inventory and documenting problems was countered. Key changes were identified (Table 1) and changes implemented, resulting in improved completion by those who agreed to complete the survey from 75% in Phase 1 to 94% in Phase 4. Satisfaction ratings also improved over time. Future research that integrates this computerized screening inventory with an EHR and assesses clinician acceptability and feasibility is needed. In addition, rigorous testing of this or similar computerized screeners in other settings, including outpatient, inpatient, and specialty medical settings, and in multilingual populations is needed to replicate and extend these findings.

Authors' Contributions

EDB was responsible for study design, execution, analysis, interpretation, and article preparation; AF, ZZ, and SS were responsible for study execution, analysis, interpretation, article preparation; BH was responsible for study execution, interpretation, and article preparation; and GC contributed to study interpretation and article preparation.

Conflicts of Interest

Dr Boudreaux has a potential financial conflict of interest related to this study. Vecna Technologies and the University of Massachusetts Medical School jointly own intellectual property incorporated in the computerized screening inventory. The organizations retain the right to commercialize the computerized screening inventory for financial gain. Revenue generated by the University of Massachusetts Medical School is shared with Dr Boudreaux, the Principal Investigator, who is an employee in this university. None of the other authors have a potential conflict of interest.

Multimedia Appendix 1

Computerized screeners used in this study.a.

[PDF File (Adobe PDF File), 98KB - humanfactors_v3i1e10_app1.pdf]

Multimedia Appendix 2

Screener results (N=1280).a.

[PDF File (Adobe PDF File), 231KB - humanfactors_v3i1e10_app2.pdf]

References

1. Estabrooks PA, Boyle M, Emmons KM, Glasgow RE, Hesse BW, Kaplan RM, et al. Harmonized patient-reported data elements in the electronic health record: Supporting meaningful use by primary care action on health behaviors and key psychosocial factors. J Am Med Inform Assoc 2012;19(4):575-582 [FREE Full text] [doi: 10.1136/amiajnl-2011-000576] [Medline: 22511015]
2. Goldstein LA, Connolly GMB, Thompson SM, Scott K, Heintz L, Green P, et al. Outcome assessment via handheld computer in community mental health: Consumer satisfaction and reliability. J Behav Health Serv Res 2011 Jul;38(3):414-423 [FREE Full text] [doi: 10.1007/s11414-010-9229-4] [Medline: 21107916]
3. Graham A, Goss C, Xu S, Magid DJ, DiGuiseppi C. Effect of using different modes to administer the AUDIT-C on identification of hazardous drinking and acquiescence to trial participation among injured patients. Alcohol Alcohol 2007;42(5):423-429 [FREE Full text] [doi: 10.1093/alcalc/agl123] [Medline: 17341515]

4. Pealer LN, Weiler RM, Pigg RM, Miller D, Dorman SM. The feasibility of a web-based surveillance system to collect health risk behavior data from college students. Health Educ Behav 2001 Oct;28(5):547-559. [Medline: 11575685]
5. Nielsen J. Usability 101: Introduction to Usability. 2012. URL: http: //www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2015-12-19] [WebCite Cache ID 6ZgEsLbGH]
6. Nielsen J. Designing Web Usability : The Practice of Simplicity. Indianapolis, IN: New Riders Press; 2000.
7. Hess R, Santucci A, McTigue K, Fischer G, Kapoor W. Patient difficulty using tablet computers to screen in primary care. J Gen Intern Med 2008 Apr;23(4):476-480 [FREE Full text] [doi: 10.1007/s11606-007-0500-1] [Medline: 18373148]
8. Mullen KH, Berry DL, Zierler BK. Computerized symptom and quality-of-life assessment for patients with cancer part II: Acceptability and usability. Oncol Nurs Forum 2004 Sep;31(5):E84-E89. [doi: 10.1188/04.ONF.E84-E89] [Medline: 15378105]
9. Cox A, Illsley M, Knibb W, Lucas C, O'Driscoll M, Potter C, et al. The acceptability of e-technology to monitor and assess patient symptoms following palliative radiotherapy for lung cancer. Palliat Med 2011 Oct;25(7):675-681. [doi: 10.1177/0269216311399489] [Medline: 21474620]
10. Bernstein SL, D'Onofrio G. Public health in the emergency department: Academic Emergency Medicine consensus conference executive summary. Acad Emerg Med 2009 Nov;16(11):1037-1039. [doi: 10.1111/j.1553-2712.2009.00548.x] [Medline: 20053218]
11. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT, The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care. 2nd edition. Geneva, Switzerland: World Health Organization; 2001. URL: http://apps.who.int/iris/bitstream/10665/67205/1/WHO_MSD_MSB_01.6a.pdf [accessed 2015-12-19] [WebCite Cache ID 6aoBhwNPf]
12. Dawson DA, Smith SM, Saha TD, Rubinsky AD, Grant BF. Comparative performance of the AUDIT-C in screening for DSM-IV and DSM-5 alcohol use disorders. Drug Alcohol Depend 2012 Dec 1;126(3):384-388 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.05.029] [Medline: 22728044]
13. Payne KB, Wharrad H, Watts K. Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): A regional survey. BMC Med Inform Decis Mak 2012;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]

Abbreviations

ANOVA: analysis of variance
AUDIT: Alcohol Use Disorders Identification Test
EHR: electronic health record
NIH: National Institutes of Health
OBSSR: Office of Behavioral and Social Sciences Research
RA: research assistant

Edited by G Eysenbach; submitted 03.07.15; peer-reviewed by M Boyle, S Raymond; comments to author 23.07.15; revised version received 20.08.15; accepted 22.09.15; published 09.03.16.

Please cite as:
Boudreaux ED, Fischer AC, Haskins BL, Saeed Zafar Z, Chen G, Chinai SA
Implementation of a Computerized Screening Inventory: Improved Usability Through Iterative Testing and Modification
JMIR Human Factors 2016;3(1):e10
URL: http://humanfactors.jmir.org/2016/1/e10/
doi:10.2196/humanfactors.4896
PMID:27025766

Original Paper

How Does Learnability of Primary Care Resident Physicians Increase After Seven Months of Using an Electronic Health Record? A Longitudinal Study

Martina A Clarke[1*], PhD; Jeffery L Belden[2*], MD; Min Soon Kim[3,4*], PhD

[1]Department of Internal Medicine, Division of Cardiology, University of Nebraska Medical Center, Omaha, NE, United States

[2]Department of Family and Community Medicine, University of Missouri, Columbia, MO, United States

[3]Department of Health Management and Informatics, University of Missouri, Columbia, MO, United States

[4]Informatics Institute, University of Missouri, Columbia, MO, United States

[*]all authors contributed equally

Corresponding Author:
Min Soon Kim, PhD
Department of Health Management and Informatics
University of Missouri
CE728 Clinical Support & Education, DC006.00
5 Hospital Drive
Columbia, MO,
United States
Phone: 1 573 884 0115
Fax: 1 573 882 6158
Email: kimms@health.missouri.edu

Abstract

Background: Electronic health records (EHRs) with poor usability present steep learning curves for new resident physicians, who are already overwhelmed in learning a new specialty. This may lead to error-prone use of EHRs in medical practice by new resident physicians.

Objective: The study goal was to determine learnability gaps between expert and novice primary care resident physician groups by comparing performance measures when using EHRs.

Methods: We compared performance measures after two rounds of learnability tests (November 12, 2013 to December 19, 2013; February 12, 2014 to April 22, 2014). In Rounds 1 and 2, 10 novice and 6 expert physicians, and 8 novice and 4 expert physicians participated, respectively. Laboratory-based learnability tests using video analyses were conducted to analyze learnability gaps between novice and expert physicians. Physicians completed 19 tasks, using a think-aloud strategy, based on an artificial but typical patient visit note. We used quantitative performance measures (percent task success, time-on-task, mouse activities), a system usability scale (SUS), and qualitative narrative feedback during the participant debriefing session.

Results: There was a 6-percentage-point increase in novice physicians' task success rate (Round 1: 92%, 95% CI 87-99; Round 2: 98%, 95% CI 95-100) and a 7-percentage-point increase in expert physicians' task success rate (Round 1: 90%, 95% CI 83-97; Round 2: 97%, 95% CI 93-100); a 10% decrease in novice physicians' time-on-task (Round 1: 44s, 95% CI 32-62; Round 2: 40s, 95% CI 27-59) and 21% decrease in expert physicians' time-on-task (Round 1: 39s, 95% CI 29-51; Round 2: 31s, 95% CI 22-42); a 20% decrease in novice physicians mouse clicks (Round 1: 8 clicks, 95% CI 6-13; Round 2: 7 clicks, 95% CI 4-12) and 39% decrease in expert physicians' mouse clicks (Round 1: 8 clicks, 95% CI 5-11; Round 2: 3 clicks, 95% CI 1-10); a 14% increase in novice mouse movements (Round 1: 9247 pixels, 95% CI 6404-13,353; Round 2: 7991 pixels, 95% CI 5350-11,936) and 14% decrease in expert physicians' mouse movements (Round 1: 7325 pixels, 95% CI 5237-10,247; Round 2: 6329 pixels, 95% CI 4299-9317). The SUS measure of overall usability demonstrated only minimal change in the novice group (Round 1: 69, high marginal; Round 2: 68, high marginal) and no change in the expert group (74; high marginal for both rounds).

Conclusions: This study found differences in novice and expert physicians' performance, demonstrating that physicians' proficiency increased with EHR experience. Our study may serve as a guideline to improve current EHR training programs. Future directions include identifying usability issues faced by physicians when using EHRs, through a more granular task analysis to recognize subtle usability issues that would otherwise be overlooked.

(JMIR Human Factors 2016;3(1):e9) doi:10.2196/humanfactors.4601

KEYWORDS

primary care, physicians, usability, electronic health records, computerized physician order entry, user-computer interface

Introduction

Physicians' Electronic Health Records (EHR) Use

Health information technology's (HIT) functionality in clinical practice is expanding and physicians are increasingly adopting EHRs as a result of the financial incentives guaranteed by Centers for Medicare & Medicaid Services (CMS) [1]. Meaningful Use (MU) is one measure of successful adoption of EHRs as a component of the Health Information Technology for Economic and Clinical Health (HITECH) act proposed by the Office of the National Coordinator for Health Information Technology (ONC) and CMS. EHRs are "records of patient health information generated by visits in any health care delivery setting" [2]. EHRs center on the overall health of a patient beyond clinical data gathered from a single provider, and offer a more comprehensive view of a patient's care. EHRs are designed for sharing data with other health care providers such as laboratories and specialists; therefore, EHRs contain information from every clinician involved in a patient's care [3]. In a data brief in 2013, the National Center for Health Statistics (NCHS) reported that 78% of office-based physicians had adopted EHRs in their practice [4]. Presently, EHRs require a large investment of effort for users to become proficient in their use. Resident physicians were selected for this study because those who are not adequately trained in using EHRs may experience a steep learning curve when beginning their residency program [5]. In an effort to maximize physician proficiency with EHRs, hospitals and clinics provide comprehensive EHR training for their resident physicians. However, it is challenging to find sufficient time to train physicians to use new EHR systems [6-9]. Using information technology to manage the process of patient care and to communicate with patients is an essential redesign of clinical practice [10]. Some advantages expressed by EHR users of adopting an EHR consist of the following: increased adherence to guidelines in preventive care, decreased paperwork for providers, and improvement in overall quality and efficiency of patient care [11-13]. Nonetheless, there are possible drawbacks to EHRs: financial burden, mismatch of human and machine workflow models, and productivity loss potentially caused by EHR usability issues [11,12,14-22]. Usability is described as the degree to which software can by employed by users to effectively perform a particular task in a specific content area [23]. EHRs with poor usability may have a negative effect on clinicians' EHR learning experience. This could lead to increased cognitive load, medical errors, and a decline in quality of patient care [24-29]. Learnability is defined as the extent to which a system permits users to understand how to use it [30]. Learnability deals with the amount of time and effort needed for a user to develop proficiency with a system over time and after multiple use [31]. In the literature, while there are variations in defining usability and learnability [32-34], definitions of learnability are strongly correlated with usability and proficiency [33,35,36]. Allowing physicians to efficiently accomplish clinical tasks within the EHR may ease time constraints experienced by physicians during patient visits.

According to an EHR user satisfaction survey completed in 2012 by 3088 family physicians, approximately 62% of survey respondents were not satisfied with many of the best-known EHR systems, and EHR vendor support and training were the areas with lowest satisfaction ratings [37]. Multiple studies on successful EHR implementations have stressed the usefulness of training in the implementation process [7,9,38-47]. A survey by Aaronson et al [44] concerning EHR use in 219 family practice residency programs indicated that resident physicians' EHR training may have an impact not only on perceived ease of use of EHR systems, but also on the use of EHR systems in their practices after residency.

Prior EHR Usability Evaluation Studies

Previous studies have shown the importance of usability evaluation in the EHR adoption and implementation process. Current best practices promote the use of cognitive approaches to examine human-computer interactions in EHR systems [2,48-50]. Khajouei and Jasper performed a systematic review examining the impact of the design aspects of medication systems in computerized physician order entry systems (CPOE) (usually integrated in EHRs) on usability. They found that proper CPOE system design is fundamental to promoting physicians' adoption and diminishing medication errors [51]. Multiple studies have used heuristic evaluation as a method to identify usability issues in health information technology. Chan et al evaluated the usability of a CPOE order set system using heuristic evaluation and discovered 92 unique heuristic violations across 10 heuristic principles [52]. Harrington and Porch investigated an EHR's usability and identified 14 usability heuristics that were violated 346 times in the intensive care unit clinical documentation [53]. Li et al evaluated clinical decision support with simulated usability testing using a think-aloud protocol, and found that 90% of negative comments from users were concerning navigation and workflow issues [54]. In a study at an urban medical center in New York, Kushniruk et al probed the association between usability tests and training of a commercial EHR system. About 1 month after in-class training, laboratory-based usability testing containing 22 sets of scenario-based tasks was conducted. Usability issues were identified as physicians completed their tasks, leading to numerous areas of potential improvement for system learnability and usability.

Objective

EHRs with poor usability present steep learning curves for new resident physicians, who are already overwhelmed in learning a new specialty. This may lead to error-prone use of EHRs in medical practice by new resident physicians. Identifying and addressing early barriers in the learning environment can help improve the overall capacity of new physicians and save costs for organizations. The objective of this study was to determine

the difference in learnability by comparing changes in performance measures between expert and novice primary care physicians 3 and 7 months after 2 rounds of learnability tests (Round 1: November 12, 2013 to December 19, 2013; round 2: February 12, 2014 to April 22, 2014). We analyzed learnability by addressing 2 specific research questions: (1) Do performance measures of expert and novice physicians improve after 3 and 7 months of EHR experience? and (2) Does the learnability gap between novice and expert physician groups change after 7 months of EHR experience?

Methods

Study Design

To determine learnability gaps between expert and novice physicians when using EHRs, data were collected through learnability testing using Morae video analysis software (TechSmith). Twelve family medicine and 4 internal medicine resident physicians performed 19 artificial, scenario-based tasks in a laboratory setting. Four types of quantitative performance measures, a system usability scale (SUS), a survey instrument [55], and a qualitative debriefing session with participants were employed. This study was approved by the University of Missouri Health Sciences Institutional Review Board.

Organizational Setting

This study took place at the University of Missouri Health System (UMHS), which is a 536-bed, tertiary-care academic medical hospital located in Columbia, Missouri. In 2012, UMHS had approximately 553,300 clinic visits and employed more than 70 primary care physicians. The Department of Family and Community Medicine (FCM) runs 6 clinics, while the Department of Internal Medicine (IM) oversees 2 primary care clinics [56]. The Healthcare Information and Management Systems Society (HIMSS), a non-profit organization that scores how effectively hospitals employ electronic medical record (EMR) applications, assigned UMHS a rating of Stage 7 with respect to the EMR Adoption Model [57]. In other words, UMHS has adopted electronic patient charts, examined clinical data through data warehousing, and shares health information electronically with authorized health care bodies [58]. The CPOE within the EHR permits physicians to safely and electronically access and place lab and medication orders for patients, and transfer orders directly to departments that are responsible for implementing requests. UMHS' EHR database comprises all data from the university's hospitals and clinics. University of Missouri Health Care has been using a mature EHR system since 2003 from the same vendor. New users of the EHR receive 4 to 8 hours of training and also have drop-in access (or can book an appointment) to an EHR Help Room to receive help or further training. Supplemental online training materials such as documents, videos, and self-paced tutorials are also available. When new features are included in the EHR, illustrated instructions and explanations become available.

Participants

There is presently no evidence-based approach to measure a user's EHR experience; therefore, novice and expert physicians were distinguished based on clinical training level and number of years using the EHR. This decision was based on a discussion with an experienced physician champion (JLB). This study will examine and confirm if after 1 year of EHR use, resident physicians have gained sufficient skills to be considered an expert [59]. Thus, 10 first-year resident physicians were grouped as novice users and 6 second and third-year resident physicians were grouped as expert users. Both FCM and IM run 3-year residency programs. A convenience sampling method was used when choosing participants [60]. UMHS FCM and IM physicians were selected for the sample because, as primary care residents, they have equivalent clinical roles and duties. Based on a review of the literature, a sample of 15 to 20 participants was judged suitable for exploratory usability studies to identify major problems to correct in a product development cycle [61-63]. However, we observed data saturation in terms of usability issues at 5 participants. Participation was voluntary and subjects were compensated US $20 for their involvement in the project.

In Round 1, 10 novice physicians and 6 expert physicians participated in the study. Out of the 10 novice physicians in Round 1, 7 were from family medicine and 3 from internal medicine. Of the 10 novice physicians, 6 (60%) were male, 8 (80%) identified their race as white, 1 (10%) identified as Asian, and 1 (10%) identified as both Asian and white. The age of novice physicians ranged from 27 to 31 and the mean age was 28 years. In Round 1, 4 (40%) novice physicians had no experience with an EHR other than the one at UMHS, 2 (20%) had less than 3 months of experience, 1 (10%) had 7 months to 1 year of experience, and 3 (30%) had over 2 years of experience with an EHR other than the one at UMHS. In this study, 5 family medicine and 1 internal medicine expert physicians participated in the study. Of the 6 expert physicians, 5 (83%) were female and all (100%) identified their race as white. In this study, 2 did not provide information on their date of birth and EHR experience and were not included in the calculation of age range, mean age, and EHR experience. The age of expert physicians ranged from 30 to 33 and the mean age was 31 years. In this study, 1 (17%) expert physician had no experience with an EHR other than the one at UMHS, 1 (17%) had 7 months to 1 year of experience, and 2 (33%) had over 2 years of experience with an EHR other than the one at UMHS.

Of the 8 novice physicians and 4 expert physicians who participated in Round 1 also participated in Round 2 of the study. A total of 2 novice and 2 expert physicians who participated in Round 1 declined participation in Round 2. Conducting 2 rounds of data collection was a major strength of this study, because it allowed us to measure valid learnability. Out of the 8 novice physicians in Round 2, 5 were from family medicine and 3 from internal medicine. Of the 8 novice physicians, 5 (63%) were male, 8 (75%) identified their race as white, 1 (13%) identified as Asian, and 1 (13%) identified as both Asian and white. The age of novice physicians ranged from 27 to 30 and the mean age was 28 years. In Round 2, 3 (38%) novice physicians had no experience with an EHR other than the one at UMHS, 2 (25%) had less than 3 months of experience, 1 (13%) had 7 months to 1 year of experience, and 2 (25%) had over 2 years of experience with an EHR other than the one at UMHS. Four family medicine expert physicians participated in

the study. All 4 (100%) were female and all (100%) identified their race as white. The age of expert physicians ranged from 30 to 33 and the mean age was 31 years. In this study, 1 (25%) expert physician had no experience with an EHR other than the one at UMHS, 1 (25%) had 7 months to 1 year of experience, and 2 (50%) had over 2 years of experience with an EHR other than the one at UMHS. Because of the small sample size, we did not attempt to control for age or gender.

Scenario and Tasks

Two sets of artificial but realistic scenario-based tasks were used in the study. The tasks were created based on discussion with an experienced physician champion (JLB) and 2 chief resident physicians from both participating departments (FCM, IM). When completing Round 1 of the learnability test, resident physicians were given a scenario for a "scheduled follow-up visit after a hospitalization for pneumonia." When completing Round 2 of the learnability test, resident physicians were given a scenario for a "scheduled follow-up visit after a hospitalization for heart failure." While different, these 2 scenarios were equivalent in difficulty, workflow, and functionalities used. These scenarios were employed to assess physicians' use of the EHR with realistic inpatient and outpatient information. We included 19 tasks that are generally completed by both novice and expert primary care physicians. These tasks also met 2014 EHR certification criteria 45 CFR 170.314 for meaningful use (MU) Stage 2 [31]. The alphanumeric code located beside each task corresponds to the EHR certification criteria that satisfies meaningful use Stage 2 objectives. In order to measure learnability more effectively, we confirmed that the tasks were also practiced during EHR training required of resident physicians at the commencement of their residency. The tasks had clear objectives that physicians were able to follow without needless clinical cognitive load or ambiguity, which would have deviated from the study aim. The tasks were as follows:

1. Start a new note (§170.314[e][2]).

2. Include visit information (§170.314[e][2]).

3. Include chief complaint (§170.314[e][2]).

4. Include history of present illness (§170.314[e][2]).

5. Review current medications contained in the note (§170.314[a][6]).

6. Review problem list contained in the note (§170.314[a][5]).

7. Document new medication allergy (§170.314[a][7]).

8. Include review of systems (§170.314[e][2]).

9. Include family history (§170.314[a][13]).

10. Include physical exam (§170.314[a][4] and §170.314[e][2]).

11. Include last comprehensive metabolic panel (CMP) (§170.314[b][5]).

12. Save the note.

13. Include diagnosis (§170.314[a][5]).

14. Place order for chest X-ray (§170.314[a][1] and §170.314[e][2]).

15. Place order for basic metabolic panel (BMP) (§170.314[a][1] and §170.314[e][2].

16. Change a medication (§170.314[a][1] and §170.314[a][6].

17. Add a medication to your favorites list (§170.314[a][1].

18. Renew one of the existing medications (§170.314[a][1] and §170.314[a][6].

19. Sign the note.

Performance Measures

Learnability was evaluated using 4 quantitative performance measures. *Percent task success* was the percentage of subtasks that participants successfully completed without error. *Time-on-task* calculated how long in seconds it took each participant to complete each task. Calculation began when a participant clicked on the "start task" button and ended when the "end task" button was clicked. *Mouse clicks* computed the number of times the participant clicked on the mouse when completing a given task. *Mouse movement* calculated in pixels the distance of the navigation path by the mouse to complete a given task.

For percent task success rate, a higher value usually signifies better performance, representing participants' skill with the system. For time-on-task, mouse clicks, and mouse movements, a higher value usually indicates poorer performance [62,64,65]. As such, higher values may indicate that the participant encountered complications while using the system.

System Usability Scale

After testing, participants were asked to complete the System Usability Scale (SUS) to supplement the performance measures. The SUS is a 10-item survey measured on a Likert scale that provides fairly robust measures of subjective usability and is a widely used, validated instrument in HIT evaluation [31,55,66]. The SUS produces a single score (ranging from 0 to 100, with 100 being a perfect score [55]) that represents a composite measure of the overall usability of the system under examination. A score of 0 to 50 is considered not acceptable, 50 to 62 is low marginal, 63 to 70 is high marginal, and 70 to 100 is acceptable.

Data Collection

Two rounds of data collection were scheduled to measure learnability by comparing whether participants' performance measures (task success, time-on-task, mouse clicks, and mouse movements) improved and if participants experienced fewer usability issues with longer exposure to the system. Learnability pertains to the amount of time and effort needed for a user to develop proficiency with a system over time and after multiple use [31]. The 2 groups (novice and expert physicians) were essential for our comparison, because experts' measures were used to examine novices' improvements toward becoming an expert. Round 1 learnability data were collected between November 12, 2013 and December 19, 2013 and Round 2 data were collected between February 12, 2014 and April 22, 2014. Round 1 data collection began 3 months after novice (Year 1) resident physicians completed their initial mandatory EHR training at UMHS. Resident physicians were invited to complete

Round 2 approximately 3 months after the date they completed Round 1. Learnability testing was completed in approximately 20 minutes and conducted on a 15-inch laptop using Windows 7 operating system. To preserve consistency and reduce undesirable interruptions, the participant and facilitator were the only 2 individuals in the conference room. At the beginning of the session, participants were advised that their participation in the study was voluntary and they had the right to end the session at any time. Participants were provided with a binder that contained instructions on how to complete the task before the test began. Tasks were displayed at the top of the display as the test progressed. A think-aloud strategy was used throughout the session and audio, video, on-screen activity, and inputs from the keyboard and mouse were recorded using a Morae Recorder [67,68]. We prompted participants to talk aloud and describe what they were doing while completing the tasks. Participants completed the tasks without the assistance of the facilitator who would only intervene if there were any technical difficulties. However, there were none and the facilitator did not have to intervene. After participants completed the tasks, they completed the SUS and demographic survey. The test session concluded with a debriefing session during which participants were asked to comment on the specific tasks they found difficult. Interesting observations detected by the facilitator were discussed as well.

Data Analysis

We confirmed there were no EHR interface changes between data collection in Rounds 1 and 2 that may have influenced the study and tasks. The recorded sessions were examined using Morae Manager, a video analysis software program that was used to calculate performance measures using markers to identify difficulties and errors the participants encountered. Video analysis took approximately 1.5 hours for each 20-minute recorded session. The first step in the analysis was to review the recorded sessions and label any tasks that were unmarked during data collection. The second step was to divide each of the 19 tasks into smaller tasks to determine the task success rate and identify subtle usability challenges that we may have otherwise failed to notice. Geometric means were calculated for the performance measures with confidence intervals at 95% [69]. Performance measures have a strong tendency to be positively skewed, so geometric means were used because they provide the most accurate measure for sample sizes less than 25 [70]. The learnability comparison was a *between* comparison of 2 *within* comparisons. Therefore, we measured the difference between the novice and expert resident physician groups and the difference within novice and expert physician groups, 3 and 7 months after EHR training. Comparisons of learnability between the 2 groups were *between* comparisons. Time-on-task, mouse clicks, and mouse movements were measured while users interacted with the EHR system and performance measures were calculated automatically by the Morae Manager usability analysis software program. Percent task success was calculated by creating subtasks out of each task and then identifying each subtask the physician completed successfully. For example, for Task 8 (Include review of systems) the subtasks created to calculate the task success rate were the following: (1) go to review of systems, (2) add “no chills,” (3) add “no fever,” (4) add “fatigue,” (5) add “decreased activity,” (6) add “dry mouth,” (7) add “no dyspnea,” and (8) add “no edema.”

Results

Percent Task Success Rate

Geometric mean values of percent task success rates were compared between the 2 physician groups across 2 rounds (Table 2) [69]. There was a 6-percentage-point increase in the novice physician group’s percent task success rate between Round 1 (92%, 95% CI 87%-99%) and Round 2 (98%, 95% CI 95%-100%). Similarly, expert physicians had a 7-percentage-point increase in percent task success rate between Round 1 (90%, 95% CI 83%-97%) and Round 2 (97%, 95% CI 93%-100%). When mean task success rates were compared between the physician groups, the novice physician group had a higher task success rate than the expert physician group did for both rounds.

Table 2. Geometric mean values of performance measures for novice and expert physicians across two rounds.

Performance Measures	Round 1 Novice	Round 2 Novice	Round 1 Expert	Round 2 Expert
Task Success	92%	98%	90%	97%
Time-on-Task	44	40	39	31
Mouse Clicks	8	7	8	5
Mouse Movements	9247	7992	7325	6329

In Round 1, the novice physician group achieved a higher success rate than expert physicians for 7 tasks (2, 8, 11, 13, and 15-17), the same success rate for 7 tasks (1, 3-6, 9, and 19), and a lower success rate for 5 tasks (7, 10, 12, 14, and 18). In Round 2, the novice physician group achieved a higher success rate for 3 tasks (8, 9, and 14), the same success rate for 15 tasks (1-7, 10-13, and 16-19), and a lower success rate for Task 15.

Both novice (6%) and expert physician groups (2%) had equally low task success for Task 7 (Add a medication to your favorites list) in Round 1. However, in Round 2 all physicians in both groups successfully completed Task 7 (100%).

Time-on-Task

Geometric mean values of time-on-task (TOT) were compared between the 2 physician groups across the 2 rounds (Table 2). There was a 10% decrease in novice physicians’ time-on-task between Round 1 (44s, 95% CI 32-62) and Round 2 (40s, 95% CI 27-59). There was a 21% decrease in the expert physician group’s time-on-task between Round 1 (39s, 95% CI 29-51) and Round 2 (31s, 95% CI 22-42). When time-on-task was

compared between the physician groups, the overall novice physician group spent more time compared to the expert physician group for both rounds.

In Round 1, the novice physician group spent less time than expert physicians completing 4 out of 19 tasks (5, 11, 12, and 13), the same amount of time completing Task 17, and more time completing 14 tasks (1-4, 6-10, 14-16, 18, and 19). In Round 2, the novice physician group spent less time completing 4 out of 19 tasks (2, 6, 11, and 12), the same time completing Task 18, and more time completing 14 tasks (1, 3-5, 7-10, 13-17, and 19).

In Round 1, both physician groups had the longest time spent on Task 7 (Document new medication allergy). However, in Round 2, time on Task 7 decreased by 52% for the expert physician group (87s to 50s) and 29% for the novice physician group (133s to 95s).

Mouse Clicks

Geometric mean values of mouse clicks were compared between the 2 physician groups across the 2 rounds (Table 2). There was a 20% decrease in the novice physician group's mouse clicks between Round 1 (8 clicks, 95% CI 6-13) and Round 2 (7 clicks, 95% CI 4-12). Similarly, there was a 39% decrease in the expert physician group's mouse clicks between Round 1 (8 clicks, 95% CI 5-11) and Round 2 (5 clicks, 95% CI 1-10). When mouse clicks were compared between the physician groups, the novice physician group completed tasks with slightly more mouse clicks than expert physicians did in both rounds.

In Round 1, the novice physician group achieved lower mouse clicks than the expert physician group for 7 tasks (4, 6, 8, 11, 13, 17, and 19), higher mouse clicks for 9 tasks (1, 5, 7, 9 10, 12, and 14 – 16), and a comparable number of clicks for 3 tasks (2, 3, and 18). In Round 2, novice physicians used less mouse clicks when completing 6 tasks (8, 10, 11, 13, 18 and 19), the same number of clicks when completing 5 tasks (4-6, 12, and 15), and more clicks completing 8 tasks (1-3, 7, 9, 14, 16, and 17).

In Round 1, both novice and expert physicians had the highest number of mouse clicks out of all tasks when completing Task 7 (Add a medication to your favorites list). However, in Round 2, the task with the highest number of mouse clicks by expert physicians changed from Task 7 to Task 15 (Place order for basic metabolic panel [BMP]) and novice physicians had the highest mouse clicks when completing Task 14 (Place order for chest X-ray) in Round 2, compared to Task 7 in Round 1.

Mouse Movements

Geometric mean values of mouse movements (the length of the navigation path to complete a given task) were compared between the 2 physician groups across the 2 rounds. There was a 14% increase in novice physicians' mouse movements between Round 1 (9247 pixels, 95% CI 6404-13,353) and Round 2 (7992 pixels, 95% CI 5350-11,936). There was also a 14% decrease in expert physicians' mouse movements between Round 1 (7325 pixels, 95% CI 5237-10,247]) and Round 2 (6329 pixels, 95% CI 4299-9317). When mouse movements were compared between the physician groups, the novice physician group showed slightly longer mouse movements than expert physicians did across the 19 tasks in both rounds.

In Round 1, the novice physicians showed longer mouse movements for 15 of 19 tasks (1-4, 6-12, 14-16, and 18), and shorter mouse movements for 4 tasks (5, 13, 17, and 19). In Round 2, novice physicians used shorter mouse movements in completing 8 out of 19 tasks (2, 4, 6, 11-13, 18, and 19) and used longer movements completing 11 tasks (1, 3, 5, 7-10, and 14-17).

In Round 1, novice physicians had the longest mouse movements out of all tasks when completing Task 7 (Add a medication to your favorites list) and expert physicians had the longest mouse movements when completing Task 13 (Include diagnosis). In Round 2, the task with the longest mouse movements by novice physicians was Task 14 (Place order for chest X-ray) compared to Task 7 in Round 1 and expert physicians had the longest mouse movements when completing Task 15 (Place order for basic metabolic panel [BMP]).

System Usability Scale

In Round 1, 5 out of 6 expert physicians and all 10 novice physicians completed the SUS. In Round 2, all 4 expert physicians and all 9 novice physicians completed the SUS. The SUS illustrated that novice physicians ranked the system's usability at a mean of 69 (high marginal) in Round 1 compared to 68 (high marginal) in Round 2. Experts rated the system's usability at a mean of 74 (acceptable) in both rounds. A novice physician and 2 expert physicians had a score of 50 (not acceptable) or below. These results may indicate that expert users who have achieved a certain level of proficiency may be more confident using the EHR than novice users. A debriefing session confirmed the overall learnability test experience but did not reveal specific learnability issues. After analyzing the recording, however, it was clear that physicians encountered some difficulties when completing the tasks.

Usability Themes

Because of space limitations, a second manuscript is in preparation with a full review of the usability themes. Sub-task analysis was instrumental in identifying multiple usability concerns. We identified 31 common and 4 unique usability issues between the 2 physician groups across 2 rounds. Themes were created by analyzing and combining usability issues to form an overarching theme [71]. Five themes emerged during analysis: 6 usability issues were related to inconsistencies, 9 issues concerning user interface issues, 6 issues in relation to structured data issues, 7 ambiguous terminology issues, and 6 issues in regards to workarounds. An example of an inconsistency issue was illogical ordering of lists in Task 17 (Add a medication to your favorites list), such that the medication list could not be sorted alphabetically when imported into a patient's visit note. This may frustrate physicians when they cannot discern how to sort the medication list. An example of a user interface issue was the long note template list physicians had to navigate when they completed Task 1 (Start a new note). A lengthy list of different templates was chosen from when creating a note and the templates were not specialty specific, such that searching through the template list and

choosing a desired template was time consuming and caused extra cognitive load. An example of a structured data issue was a lack of distinction between columns in Task 9 (Include Family History). In this task, the blue or white columns (indicating negative vs positive findings) for family members were unlabeled, such that physicians were unsure how to mark a family history item "positive." An example of an ambiguous terminology issue was multiple fields having the same functionality. When completing Tasks 14 and 15, there was no clear difference between the drop-down menu labeled "Requested Start Date," the drop-down menu labeled "Requested Time Frame," and the radio button labeled "Future Order." This could cause future lab tests not to be ordered properly, such that lab tests may not be completed at the right time and patients may have to get the test redone, which adds additional cost for the patient. An example of a workaround was unawareness of functions. When completing Task 13 (Include diagnosis), physicians were not able to move "hypertension" from the problem list to the current diagnosis list, so they re-added "hypertension" as a new problem, which took additional time.

Discussion

Principal Findings

Our findings show that there were mixed changes in performance measures and expert physicians were more proficient than novice physicians on all four performance measures.

Relation to Prior Studies

In our study, differences were found between expert and novice physicians' performance measures across Round 1 and Round 2. A study by Kjeldskov, Skov, and Stage [72] identifying usability problems encountered by novice and expert nurses examined whether or not usability issues disappeared over time. In this study, 7 nurses completed 14 and 30 hours of training prior to the first evaluation that included 7 tasks and subtasks centered on the core purpose of the system. The same nurses completed the same 7 tasks after 15 months of daily use of the system. All expert subjects solved all 7 tasks either completely or partially while only 2 novice subjects solved all tasks (P=.01). No statistically significant difference between novice and expert nurses was found when considering only completely solved tasks (P=.08). Our study did not report P values due to the small sample size; however, we observed overall improvement in performance measures for both novice and expert physician groups across 2 rounds. The contradictory results from this study and the study by Kjeldskov, Skov, and Stage, suggest that further research is necessary to draw more definite conclusions about task success between novice and expert physicians.

Alternatively, a study by Lewis et al measured performance of novice health sciences students and a predictive model of skilled human performance when performing EHR tasks using a touchscreen. Novice participants were adults with no prior experience using an EHR touchscreen interface using CogTool. CogTool is an open-source user-interface prototyping tool that uses a human performance model to automatically evaluate how efficiently a skilled user can complete a task. Participants completed 31 tasks commonly performed by nurses and patient registration clerks in an Anti-Retroviral Therapy clinic. The mean novice performance time for all tasks was significantly slower than predictions of skilled use (P<.00) [73]. Although novice EHR users completed touchscreen tasks slower than a skilled user, they were able to execute some tasks at a skilled level within the first hour of system use. Our study also found novice physicians completing tasks slower than expert physicians, although they decreased their time-on-task by 10% in Round 2. However, our study is different from Lewis et al in that we used human expert physicians instead of a predictive model, which gives a more realistic comparison between novice and expert users. The common findings between this study and those of Lewis et al suggest that physicians become efficient as EHR experience increases, in relation to task completion time, because physicians may become familiar with the system.

Physicians' perceptions of the usability of a system may have relations to learnability; that is, physicians may find the system more user-friendly (usability) if the amount of time and effort needed to develop proficiency with the system is shorter (learnability). In our study, the SUS, which measures overall usability, illustrated that there was only a slight change in novice (Round 1: 69 [high marginal], Round 2: 68 [high marginal]) and expert (Round 1: 74 [high marginal], Round 2: 74 [high marginal]) physicians' rankings of the system's usability. In a study by Haarbrandt et al, primary care providers gave a SUS rating of 70.7 (marginally acceptable) when asked about their perception of a health information exchange system, which was similar to the physicians' scores in our study. Expert and novice participants found the graphical user interface easy to use; however, they only rated the system as acceptable [74]. Kim et al [62] measured usability gaps in emergency department (ED) nurses, and found that novice ED nurses were not satisfied with their system (43 [unacceptable] to 55 [low marginal]) in comparison to expert nurses who were satisfied (75 to 81 [good to excellent]), which was different from our study's result. The varying SUS scores from the studies mentioned suggest that physicians with more experience using an EHR are more likely to give the system higher SUS scores. Contrary to the assumption that SUS produce reliable scores, there are mixed results that SUS scores clearly associate with performance measures. For example, Kim et al showed very low correlations between performance measures and SUS Scores, indicating that care needs to be taken when interpreting usability data and comprehensive rather than single measures are necessary.

Study Limitations

This study had several limitations in terms of the methodology. First, it involved a small sample of physicians; therefore, the sample size may not have been sufficient to obtain statistical significance when reporting quantitative results of learnability. However, the sample size was sufficient when identifying usability issues experienced by participants when interacting with the EHR system. This study was conducted at a health care institution where only 1 EHR system was used and may not be representative of all primary care practice. As such, the study's findings may have limited generalizability to other ambulatory clinic settings, due to different types of EHR applications and physician practice settings. However, the EHR platform

employed in this study is one of the top commercial products with significant market share. Based on data from Office of the National Coordinator for Health Information Technology, Cerner was reported as the primary EHR Vendor by 20% of hospitals participating in the CMS EHR incentive programs, making it the second most implemented EHR in hospitals [75]. Second, a limited number of clinical tasks were used in the learnability test and may not have encompassed other tasks completed by physicians in other clinical scenarios. However, these tasks included realistic inpatient and outpatient tasks that resident physicians would usually complete in a clinical scenario. Third, this study was conducted in a laboratory setting, which did not take into account common distractions physicians may experience during a clinical encounter. Nonetheless, laboratory-based learnability tests allow for flexibility in questioning and give room for more in-depth probing. Direct observation in laboratory learnability testing also allows for interaction between participant and facilitator. Although this study contained some methodological limitations, we believe it to be a well-controlled study that used a rigorous evaluation method with validated performance measures that are widely accepted in HIT evaluation. In addition, the clear instructions allowed physician participants to complete the required tasks without excessive cognitive load.

Conclusion

Overall, this study identified varying degrees of learnability gaps between expert and novice physician groups that may impede the use of EHRs. Our results suggest that longer experience with an EHR may not be equivalent to being an expert or proficient in its use. The physicians' interactions with the EHR can be communicated to EHR vendors, to assist in improving the user interface for effective use by physicians. This study may also assist in the design of EHR education and training programs by highlighting the areas (ie, tasks and related features and functionalities) of difficulty that resident physicians face. Resident physicians in primary care are offered extensive EHR training by their institutions. However, it is a great challenge for busy physicians to find time for training. Furthermore, it is an arduous task attempting to meet the needs of users and provide hands-on, on-site support [7], and evidence-based guidelines for training resident physicians effectively on how to use EHRs for patient care are scarce [76]. Thus, our study may also serve as a guideline to potentially improve EHR training programs, which may increase physicians' performance, by improving competency when using the system.

Conflicts of Interest

None declared.

References

1. Centers for Medicare & Medicaid Services. Medicare and Medicaid EHR Incentive Program Basics URL: https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/basics.html [accessed 2015-10-14] [WebCite Cache ID 6cHEh1HVi]
2. Mollon B, Chong J, Holbrook AM, Sung M, Thabane L, Foster G. Features predicting the success of computerized decision support for prescribing: A systematic review of randomized controlled trials. BMC Med Inform Decis Mak 2009;9:11 [FREE Full text] [doi: 10.1186/1472-6947-9-11] [Medline: 19210782]
3. Garrett P, Seidman J. Electronic Health & Medical Records 2011. EMR vs EHR - What is the Difference? URL: http://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/emr-vs-ehr-difference/ [accessed 2015-10-14] [WebCite Cache ID 6cHEoE9nt]
4. Hsiao C, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2012. NCHS Data Brief 2012 Dec(111):1-8 [FREE Full text] [Medline: 23384787]
5. Yoon-Flannery K, Zandieh SO, Kuperman GJ, Langsam DJ, Hyman D, Kaushal R. A qualitative analysis of an electronic health record (EHR) implementation in an academic ambulatory setting. Inform Prim Care 2008;16(4):277-284. [Medline: 19192329]
6. Carr DM. A team approach to EHR implementation and maintenance. Nurs Manage. Suppl 5 2004;35:24-16. [Medline: 15636539]
7. Terry AL, Thorpe CF, Giles G, Brown JB, Harris SB, Reid GJ, et al. Implementing electronic health records: Key factors in primary care. Can Fam Physician 2008 May;54(5):730-736 [FREE Full text] [Medline: 18474707]
8. Lorenzi N, Kouroubali A, Detmer D, Bloomrosen M. How to successfully select and implement electronic health records (EHR) in small ambulatory practice settings. BMC Med Inform Decis Mak 2009;9:15 [FREE Full text] [doi: 10.1186/1472-6947-9-15] [Medline: 19236705]
9. Whittaker AA, Aufdenkamp M, Tinley S. Barriers and facilitators to electronic documentation in a rural hospital. J Nurs Scholarsh 2009;41(3):293-300. [doi: 10.1111/j.1547-5069.2009.01278.x] [Medline: 19723278]
10. Morrison I, Smith R. Hamster health care. BMJ 2000;321(7276):1541-1542 [FREE Full text] [Medline: 11124164]
11. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]

12. Miller RH, West C, Brown TM, Sim I, Ganchoff C. The value of electronic health records in solo or small group practices. Health Aff (Millwood) 2005;24(5):1127-1137 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1127] [Medline: 16162555]
13. Shekelle PG, Morton SC, Keeler EB. Costs and benefits of health information technology. Evid Rep Technol Assess (Full Rep) 2006 Apr(132):1-71. [Medline: 17627328]
14. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011;4:47-55 [FREE Full text] [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
15. Goldzweig C, Towfigh A, Maglione M, Shekelle P. Costs and benefits of health information technology: New trends from the literature. Health Aff (Millwood) 2009;28(2):w282-w293 [FREE Full text] [doi: 10.1377/hlthaff.28.2.w282] [Medline: 19174390]
16. Grabenbauer L, Fraser R, McClay J, Woelfl N, Thompson CB, Cambell J, et al. Adoption of electronic health records: A qualitative study of academic and private physicians and health administrators. Appl Clin Inform 2011;2(2):165-176 [FREE Full text] [doi: 10.4338/ACI-2011-01-RA-0003] [Medline: 23616868]
17. Reynolds CJ, Wyatt JC. Open source, open standards, and health care information systems. J Med Internet Res 2011;13(1):e24 [FREE Full text] [doi: 10.2196/jmir.1521] [Medline: 21447469]
18. 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]
19. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the “triple aim” of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856. [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
20. Ratwani R, Fairbanks R, Hettinger A, Benda N. Electronic health record usability: Analysis of the user-centered design processes of eleven electronic health record vendors. J Am Med Inform Assoc 2015 Nov;22(6):1179-1182. [doi: 10.1093/jamia/ocv050] [Medline: 26049532]
21. Hübner U. What are complex eHealth innovations and how do you measure them? Position paper. Methods Inf Med 2015;54(4):319-327. [doi: 10.3414/ME14-05-0001] [Medline: 25510406]
22. Berlin J. Better bridges, better systems. Tex Med 2015 Sep;111(9):39-43. [Medline: 26360338]
23. ISO/TC 159/SC 4 - Ergonomics of human-system interaction. Geneva, Switzerland: International Organization for Standardization; 1998. ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part 11: Guidance on Usability URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883 [accessed 2015-12-22] [WebCite Cache ID 6dyHXugKo]
24. Clarke MA, Steege LM, Moore JL, Koopman RJ, Belden JL, Kim MS. Determining primary care physician information needs to inform ambulatory visit note display. Appl Clin Inform 2014;5(1):169-190 [FREE Full text] [doi: 10.4338/ACI-2013-08-RA-0064] [Medline: 24734131]
25. Love JS, Wright A, Simon SR, Jenter CA, Soran CS, Volk LA, et al. Are physicians' perceptions of healthcare quality and practice satisfaction affected by errors associated with electronic health record use? J Am Med Inform Assoc 2012;19(4):610-614 [FREE Full text] [doi: 10.1136/amiajnl-2011-000544] [Medline: 22199017]
26. McLane S, Turley JP. One size does not fit all: EHR clinical summary design requirements for nurses. 2012 Presented at: NI 2012: 11th International Congress on Nursing Informatics; June 23-27, 2012; Montreal, QC p. 283.
27. Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability. Int J Med Inform 2011 Oct;80(10):708-725. [doi: 10.1016/j.ijmedinf.2011.06.010] [Medline: 21784701]
28. Clarke MA, Steege LM, Moore JL, Belden JL, Koopman RJ, Kim MS. Addressing human computer interaction issues of electronic health record in clinical encounters. In: Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience.: Springer Berlin Heidelberg; 2013 Jan 01 Presented at: Second International Conference on Design, User Experience, and Usability; July 21-26, 2013; Las Vegas, NV p. 381-390. [doi: 10.1007/978-3-642-39241-2_42]
29. Clarke MA, Steege LM, Moore JL, Koopman RJ, Belden JL, Kim MS. Determining primary care physician information needs to inform ambulatory visit note display. Appl Clin Inform 2014;5(1):169-190 [FREE Full text] [doi: 10.4338/ACI-2013-08-RA-0064] [Medline: 24734131]
30. ISO/IEC JTC 1/SC 7 - Software and systems engineering. Geneva, Switzerland: International Organization for Standardization; 2011. ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=35733 [accessed 2015-12-22] [WebCite Cache ID 6dyHUB6HV]
31. Tullis T, Albert W. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Amsterdam: Elsevier/Morgan Kaufmann; 2008.
32. ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. Geneva: International Organization for Standardization; 2011:34.
33. Elliott G, Jones E, Barker P. A grounded theory approach to modelling learnability of hypermedia authoring tools. Interacting Comput 2002 Oct;14(5):547-574. [doi: 10.1016/S0953-5438(02)00021-8]
34. Nielsen J. Usability engineering. San Francisco, Calif: Morgan Kaufmann Publishers; 1993.

35. Whiteside J, Jones S, Levy P, Wixon D. User performance with command, menu, and iconic interfaces. 1985 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 1, 1985; San Francisco, CA p. 185-191. [doi: 10.1145/1165385.317490]
36. Lin H, Choong Y, Salvendy G. A proposed index of usability: A method for comparing the relative usability of different software systems. Behav Inf Technol 1997;16:267-277 [FREE Full text]
37. Edsall R, Adler K. The 2012 EHR User Satisfaction Survey: Responses from 3,088 family physicians. Fam Pract Manag 2012;19(6):23-30 [FREE Full text] [Medline: 23317129]
38. Anderson L, Stafford C. The “big bang” implementation: Not for the faint of heart. Comput Nurs 2002;20(1):14-20; quiz 20. [Medline: 11838384]
39. Ash J, Bates D. Factors and forces affecting EHR system adoption: Report of a 2004 ACMI discussion. J Am Med Inform Assoc 2005;12(1):8-12 [FREE Full text] [doi: 10.1197/jamia.M1684] [Medline: 15492027]
40. Brokel JM, Harrison MI. Redesigning care processes using an electronic health record: A system's experience. Jt Comm J Qual Patient Saf 2009 Feb;35(2):82-92. [Medline: 19241728]
41. McAlearney AS, Song PH, Robbins J, Hirsch A, Jorina M, Kowalczyk N, et al. Moving from good to great in ambulatory electronic health record implementation. J Healthc Qual 2010;32(5):41-50. [doi: 10.1111/j.1945-1474.2010.00107.x] [Medline: 20854358]
42. TEKsystems. Hanover, MD; 2013 Jun 10. EHR Implementation Survey: Proactive Consideration and Planning Lead to Successful EHR Implementation URL: http://www.teksystems.com/resources/thought-leadership/it-industry-trends/ehr-implementation-survey [accessed 2015-12-21] [WebCite Cache ID 6dyDLCohv]
43. Yan H, Gardner R, Baier R. Beyond the focus group: Understanding physicians' barriers to electronic medical records. Jt Comm J Qual Patient Saf 2012 Apr;38(4):184-191. [Medline: 22533131]
44. Aaronson JW, Murphy-Cullen CL, Chop WM, Frey RD. Electronic medical records: The family practice resident perspective. Fam Med 2001 Feb;33(2):128-132. [Medline: 11271741]
45. Keenan CR, Nguyen HH, Srinivasan M. Electronic medical records and their impact on resident and medical student education. Acad Psychiatry 2006;30(6):522-527. [doi: 10.1176/appi.ap.30.6.522] [Medline: 17139024]
46. Terry AL, Giles G, Brown JB, Thind A, Stewart M. Adoption of electronic medical records in family practice: The providers' perspective. Fam Med 2009;41(7):508-512 [FREE Full text] [Medline: 19582637]
47. Hammoud MM, Margo K, Christner JG, Fisher J, Fischer SH, Pangaro LN. Opportunities and challenges in integrating electronic health records into undergraduate medical education: A national survey of clerkship directors. Teach Learn Med 2012;24(3):219-224. [doi: 10.1080/10401334.2012.692267] [Medline: 22775785]
48. Saleem JJ, Patterson ES, Militello L, Anders S, Falciglia M, Wissman JA, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. J Am Med Inform Assoc 2007;14(5):632-640 [FREE Full text] [doi: 10.1197/jamia.M2163] [Medline: 17600106]
49. Patterson ES, Doebbeling BN, Fung CH, Militello L, Anders S, Asch SM. Identifying barriers to the effective use of clinical reminders: Bootstrapping multiple methods. J Biomed Inform 2005 Jun;38(3):189-199 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.015] [Medline: 15896692]
50. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
51. Khajouei R, Jaspers MW. The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: A systematic review. Methods Inf Med 2010;49(1):3-19. [doi: 10.3414/ME0630] [Medline: 19582333]
52. Chan J, Shojania KG, Easty AC, Etchells EE. Usability evaluation of order sets in a computerised provider order entry system. BMJ Qual Saf 2011 Nov;20(11):932-940. [doi: 10.1136/bmjqs.2010.050021] [Medline: 21807652]
53. Harrington L, Porch L, Acosta K, Wilkens K. Realizing electronic medical record benefits: An easy-to-do usability study. J Nurs Adm 2011;41(7-8):331-335. [doi: 10.1097/NNA.0b013e3182250b23] [Medline: 21799365]
54. 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]
55. Brooke J. SUS-A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996:189-194.
56. MU 2011 Annual Report. MU Healthcare 2011 URL: http://bluetoad.com/publication/?i=106794&pre=1 [accessed 2012-04-15] [WebCite Cache ID 6cHMQLgcH]
57. University Of Missouri Health Care Achieves Highest Level of Electronic Medical Record Adoption. Columbia, MO: University of Missouri Health System; 2013. University of Missouri Health Care News Releases URL: http://www.muhealth.org/news/2012/university-of-missouri-health-care-achieves-highest-level-of-ele/ [WebCite Cache ID 6epZqT04j]
58. HIMSS Analytics. U.S. EMR Adoption Model Trends URL: https://app.himssanalytics.org/docs/HA_EMRAM_Overview_ENG.pdf [accessed 2015-10-14] [WebCite Cache ID 6epZvOGVH]
59. 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]

60. Battaglia M. Convenience sampling. In: Lavrakas P, editor. Encyclopedia of Survey Research Methods. Thousand Oaks, CA: Sage Publications, Inc; 2008.
61. Barnum C. The 'magic number 5': Is it enough for web testing? Inf Des J 2003 Jan 01;11(3):160-170. [doi: 10.1075/idj.11.2.08bar]
62. Kim MS, Shapiro JS, Genes N, Aguilar MV, Mohrer D, Baumlin K, et al. A pilot study on usability analysis of emergency department information system by nurses. Appl Clin Inform 2012;3(1):135-153 [FREE Full text] [doi: 10.4338/ACI-2011-11-RA-0065] [Medline: 23616905]
63. Lowry S, Quinn M, Ramaiah M, Schumacher R, Patterson E, North R, et al. National Institute of Standards and Technology (NIST). 2012. Technical Evaluation, Testing, and Validation of the Usability of Electronic Health Records URL: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909701 [accessed 2015-12-22] [WebCite Cache ID 6dyG8SEGT]
64. Khajouei R, Peek N, Wierenga PC, Kersten MJ, Jaspers MW. Effect of predefined order sets and usability problems on efficiency of computerized medication ordering. Int J Med Inform 2010 Oct;79(10):690-698. [doi: 10.1016/j.ijmedinf.2010.08.001] [Medline: 20833104]
65. Koopman RJ, Kochendorfer KM, Moore JL, Mehr DR, Wakefield DS, Yadamsuren B, et al. A diabetes dashboard and physician efficiency and accuracy in accessing data needed for high-quality diabetes care. Ann Fam Med 2011;9(5):398-405 [FREE Full text] [doi: 10.1370/afm.1286] [Medline: 21911758]
66. Lewis J, Sauro J. The Factor Structure of the System Usability Scale. In: Human Centered Design: First International Conference.: Springer Berlin Heidelberg; 2009 Presented at: HCI International 2009; July 19-24, 2009; San Diego, CA p. 94-103. [doi: 10.1007/978-3-642-02806-9_12]
67. Van Someren MW, Barnard Y, Sandberg JAC. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. London: Academic Press; 1994.
68. 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 Human Factors 2015 Sep 10;2(2):e14 [FREE Full text] [doi: 10.2196/humanfactors.4537]
69. Cordes R. The effects of running fewer subjects on timeâ€ onâ€ task measures. Int J Hum-Comput Interact 1993 Oct;5(4):393-403 [FREE Full text] [doi: 10.1080/10447319309526076]
70. Sauro J, Lewis J. Average Task Times in Usability Tests: What to Report? 2010 Presented at: Conference in Human Factors in Computing Systems (CHI 2010); April 10-15, 2010; Atlanta, GA p. 2347-2350.
71. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
72. Kjeldskov J, Skov MB, Stage J. A longitudinal study of usability in health care: Does time heal? Int J Med Inform 2010 Jun;79(6):e135-e143. [doi: 10.1016/j.ijmedinf.2008.07.008] [Medline: 18757234]
73. Lewis ZL, Douglas GP, Monaco V, Crowley RS. Touchscreen task efficiency and learnability in an electronic medical record at the point-of-care. Stud Health Technol Inform 2010;160(Pt 1):101-105. [Medline: 20841658]
74. Haarbrandt B, Schwartze J, Gusew N, Seidel C, Haux R. Primary care providers' acceptance of health information exchange utilizing IHE XDS. Stud Health Technol Inform 2013;190:106-108. [Medline: 23823391]
75. Electronic Health Record Vendors Reported by Hospitals Participating in the CMS EHR Incentive Programs. Health IT Quick-Stat #29. 2015. URL: http://dashboard.healthit.gov/quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Hospitals.php [accessed 2015-10-01] [WebCite Cache ID 6cHLx4Doj]
76. Peled JU, Sagher O, Morrow JB, Dobbie AE. Do electronic health records help or hinder medical education? PLoS Med 2009 May 5;6(5):e1000069 [FREE Full text] [doi: 10.1371/journal.pmed.1000069] [Medline: 19434294]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
CPOE: computerized physician order entry systems
ED: emergency department
EHRs: electronic health records
FCM: Department of Family and Community Medicine
HIMSS: Healthcare Information and Management Systems Society
HIT: health information technology
HITECH: Health Information Technology for Economic and Clinical Health (act)
IM: Department of Internal Medicine
MU: meaningful use
NCHS: National Center for Health Statistics
ONC: Office of the National Coordinator for Health Information Technology
SUS: system usability scale
TOT: time-on-task
UMHS: University of Missouri Health System

Edited by G Eysenbach; submitted 01.05.15; peer-reviewed by T van Mierlo, J Avis, G Stiglic; comments to author 19.08.15; revised version received 14.10.15; accepted 05.11.15; published 15.02.16.

Please cite as:
Clarke MA, Belden JL, Kim MS
How Does Learnability of Primary Care Resident Physicians Increase After Seven Months of Using an Electronic Health Record? A Longitudinal Study
JMIR Human Factors 2016;3(1):e9
URL: http://humanfactors.jmir.org/2016/1/e9/
doi:10.2196/humanfactors.4601
PMID:27025237

Original Paper

Evaluating the Usability and Perceived Impact of an Electronic Medical Record Toolkit for Atrial Fibrillation Management in Primary Care: A Mixed-Methods Study Incorporating Human Factors Design

Kim Tran[1], BSc; Kori Leblanc[1,2,3], PharmD; Alissia Valentinis[4], MD; Doug Kavanagh[5], MD; Nina Zahr[1], BDes; Noah M Ivers[6], MD, PhD

[1]OpenLab, University Health Network, Toronto, ON, Canada
[2]Department of Pharmacy, University Health Network, Toronto, ON, Canada
[3]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada
[4]Taddle Creek Family Health Team, Toronto, ON, Canada
[5]North York Family Health Team, Toronto, ON, Canada
[6]Women's College Hospital, Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Kim Tran, BSc
OpenLab
University Health Network
200 Elizabeth Street
Toronto, ON, M5G2C4
Canada
Phone: 1 6476789375
Fax: 1 4165955826
Email: kim.tran2@uhn.ca

Abstract

Background: Atrial fibrillation (AF) is a common and preventable cause of stroke. Barriers to reducing stroke risk through appropriate prescribing have been identified at the system, provider, and patient levels. To ensure a multifaceted initiative to address these barriers is effective, it is essential to incorporate user-centered design to ensure all intervention components are optimized for users.

Objective: To test the usability of an electronic medical record (EMR) toolkit for AF in primary care with the goal of further refining the intervention to meet the needs of primary care clinicians.

Methods: An EMR-based toolkit for AF was created and optimized through usability testing and iterative redesign incorporating a human factors approach. A mixed-methods pilot study consisting of observations, semi-structured interviews, and surveys was conducted to examine usability and perceived impact on patient care and workflow.

Results: A total of 14 clinicians (13 family physicians and 1 nurse practitioner) participated in the study. Nine iterations of the toolkit were created in response to feedback from clinicians and the research team; interface-related changes were made, additional AF-related resources were added, and functionality issues were fixed to make the toolkit more effective. After improvements were made, clinicians expressed that the toolkit improved accessibility to AF-related information and resources, served as a reminder for guideline-concordant AF management, and was easy to use. Most clinicians intended to continue using the toolkit for patient care. With respect to impact on care, clinicians believed the toolkit increased the thoroughness of their assessments for patients with AF and improved the quality of AF-related care received by their patients.

Conclusions: The positive feedback surrounding the EMR toolkit for AF and its perceived impact on patient care can be attributed to the adoption of a user-centered design that merged clinically important information about AF management with user needs. This study demonstrates the utility of a human factors approach to piloting and refining an intervention prior to wide-scale implementation.

(JMIR Human Factors 2016;3(1):e7) doi:10.2196/humanfactors.4289

KEYWORDS

primary health care, atrial fibrillation, electronic health records, mixed-methods research, evidence-based medicine

Introduction

Atrial fibrillation (AF) is a common and preventable cause of stroke [1]. The prevalence of AF is approximately 1% overall; however, it accounts for 15% of all strokes and 33% of all strokes in the elderly [2]. Such strokes result in permanent disability in 60% and death in 20% of individuals [3]. Medications can effectively reduce the risk of stroke. Unfortunately, although evidence has long been available that many AF-related strokes are preventable with proper therapy, the proportion of eligible patients receiving appropriate stroke prevention therapy remains stubbornly low. The 2012 Canadian Cardiovascular Society guidelines for AF emphasize that the vast majority of patients with AF would likely benefit from anticoagulation to reduce risk of stroke [4]. However, studies have found that many patients at high risk of stroke are not receiving anticoagulation.

Barriers to appropriate stroke prevention therapy may exist at the system, physician, and patient level. At the system level, primary care clinics were found to have inadequate coordination with laboratories, ineffective INR tracking systems, and inefficient use of reminders [5]. Physicians tend to overestimate the risk of bleeding associated with anticoagulation, especially in the elderly, even though guidelines state that the benefits of anticoagulation outweigh the risks for most patients over 65 years of age [6-8]. In contrast, patients were found to place more value on avoidance of stroke than avoidance of bleeding [9]. In the context of infrequent use of formal risk assessment tools and underutilization of anticoagulation, it is plausible that tools supporting evidence-informed, shared decision-making processes with patients may lead to increased utilization of anticoagulation [10,11].

Electronic medical record (EMR) interventions have been described as instrumental for chronic disease management. Useful aspects of these interventions include decision support such as reminders for patient care, clinical monitoring through large-scale surveillance and data aggregation, and disease-specific encounter templates [12,13]. These EMR-based interventions have been shown to improve quality of care, improve efficiency, and decrease health services utilization [12-14].

We developed an EMR-based toolkit of quality improvement strategies to aid atrial fibrillation management in primary care, to improve the proportion of patients receiving guideline-concordant stroke prevention therapy. The objective of this pilot study was to test the usability of the EMR toolkit for AF in primary care with the goal of further refining the intervention to meet the needs of primary care clinicians. A human factors approach was utilized in the development of the toolkit in an effort to optimize the uptake of the toolkit, which in turn, would have the potential to increase the proportion of patients receiving guideline-concordant AF care.

Methods

Study Design

We conducted a pilot study using a mixed-methods approach consisting of observations, semi-structured interviews, and surveys to examine the usability of an EMR toolkit for AF and its perceived impact on care and workflow. A human factors approach, defined as "the study of the interrelationship between humans, the tools and equipment they use in the workplace, and the environment in which they work" was used to optimize usability and uptake of the toolkit [15]. The study received approval from the research ethics board at the University of Toronto.

Study Population

The study was conducted between October 2013 and July 2014. Participants were primary care clinicians at the Taddle Creek Family Health Team (TCFHT) and Women's College Hospital (WCH) Family Practice Health Centre in Toronto, Ontario, Canada. The TCFHT consists of 14 family physicians, 3 nurse practitioners, 3 nurses, 3 social workers, 2 pharmacists, and a dietician who provide primary health care to 18,900 patients. The WCH Family Practice Health Centre consists of 31 family physicians, 2 nurse practitioners, 16 nurses, 2 dieticians, 1 occupational therapist, 2 social workers, and 1 pharmacist who provide primary health care to 18,500 patients.

EMR Toolkit Development

The purpose of the EMR toolkit for AF, which links to PS Suite EMR, is to facilitate the uptake and translation of knowledge regarding guideline-based AF management in primary care and to improve the proportion of patients receiving guideline-based AF care. The toolkit was developed by an interdisciplinary team with clinical and design expertise. The clinical content was developed by 3 family physicians, a pharmacist, and a cardiologist with expertise in primary care and AF management. A designer provided human-centered design expertise and created mockups of the design and layout of the toolkit. Toolkit development occurred through an iterative process with iterations cycled back through the team prior to the commencement of formal usability testing. The EMR toolkit incorporated evidence-based recommendations for AF management specified in the 2012 Canadian Cardiovascular Society AF guidelines. The final version consisted of a toolbar embedded into the electronic medical charts of patients with AF that included the following tools: (1) initial assessment form; (2) follow-up visit form; (3) stroke and bleeding risk calculator, which included guideline-concordant recommendations for treatment; (4) provider resources; and (5) patient resources. The outputs from Tools 1 to 3 are embedded into patients' electronic medical charts. The provider resources consisted of three documents: (1) AFib in One Page, (2) Comparison of Anticoagulants, and (3) Anticoagulation Dosing Table. The patient resources consisted of educational documents providing information on atrial fibrillation, what to do if a patient

experiences an AF episode, treatment options, decreasing stroke risk, available anticoagulants, and cardioversion.

Usability Testing

The EMR toolkit underwent usability testing using a qualitative approach consisting of observations of primary care clinicians' interactions with the EMR toolkit for AF and semi-structured interviews with family physicians. A purposive sample of primary care clinicians from the Taddle Creek Family Health Team and Women's College Hospital Family Practice Health Centre was selected for the study.

Usability testing was conducted by members of the research team (KT and KL) and occurred in two phases. Phase 1 used the "think aloud" approach while primary care clinicians used the toolkit with test patients. Field notes employing a structured data collection form were used to record usability-related issues. Phase 2 consisted of observations of primary care clinicians' interactions with the toolkit as they conducted a visit with an actual patient with AF; observations were followed by semi-structured interviews to examine their perceptions of the toolkit. Field notes employing a structured data collection form were used to record usability-related issues during observations. Interviews were audiotaped and transcribed by a member of the research team (KT). Participant feedback was used iteratively to make improvements to the toolkit.

Inductive thematic analysis was conducted using qualitative data analysis software (NVivo 10, QSR International). Two members of the research team (KT and KL) independently coded 4 interview transcripts to identify interesting features of the data and provide a comprehensive selection of codes, which were reviewed and discussed to ensure consensus. A coding framework was developed using initial findings to guide coding of remaining interviews and field notes, with additional codes reported as they were identified. Initial codes and their coded interview extracts were organized into themes and subthemes. All themes and coded interview extracts were reviewed to ensure data within each theme formed a coherent pattern and a clear distinction between different themes was evident.

Surveys

A survey was developed by the research team to examine primary care clinicians' perceptions of the EMR toolkit for AF and its impact on patient care and management. A subset of questions was aligned with the key themes identified through usability testing. The research team reviewed the survey for face validity, comprehensiveness, and clarity. Pre-testing occurred with 2 individuals with research backgrounds who reviewed the survey for face validity and clarity. The final survey consisted of 22 questions with a combination of multiple choice, rating scale, and open-ended items. A 5-point Likert scale and 5-point scale were used to express level of agreement and degree of change, respectively.

The survey was distributed electronically to all primary care clinicians—family physicians, nurse practitioners, nurses, and pharmacists—at the TCFHT. An initial introduction email followed by 2 reminder emails was sent to all clinicians to provide them with information regarding the study and the electronic survey. To increase the response rate, paper surveys were also distributed to clinicians following the third email reminder. Descriptive statistics were generated from the survey data.

Results

Participants

A total of 14 clinicians (13 FPs and 1 NP) participated in Phase 1 or 2 of usability testing. Each clinician was observed as they used the EMR toolkit for AF with either a test patient (Phase 1-5 FPs and 1 NP) or a real patient with AF (Phase 2-8 FPs). Semi-structured interviews were conducted with all 8 FPs who participated in Phase 2 of the usability testing. The response rate for the survey was 55% (12/22). A total of 8 family physicians, 2 nurses, 1 nurse practitioner, and 1 pharmacist participated in the survey.

Phase 1 Usability Testing: Key Themes

Participants highlighted several usability-related issues that required improvements. Three overarching themes were identified, which included (1) interface-related changes, (2) additional resources for AF management, and (3) toolkit functionality issues. In total, the EMR toolkit for AF underwent 9 iterative cycles of changes based on participants' feedback to produce the final version.

Participants highlighted several interface-related issues concerning the EMR toolkit's layout, format, and language. They described a lack of an intuitive flow for the layout of the initial assessment and follow-up visit forms, confusion surrounding the format of checkboxes, and confusion around the use of abbreviations. In response, we modified the layout, format, and language to make the toolkit more user-friendly. For the layout, we optimized the structure of the assessment forms to align with the SOAP (subjective, objective, assessment, and plan) note structure, which participants described as their typical workflow when writing chart notes. Improvements were made to the formatting by creating checkboxes with yes/no options rather than checkmarks, which participants felt was better for identifying if they had missed sections. Lastly, modifications to the language of the toolkit were made to improve clarity. For example, the abbreviation OAC was changed to oral anticoagulation.

Participants expressed a desire to have additional resources for AF management. In response, we created additional resources to aid clinician decision making. A "Provider Resources" section was added to the toolbar that provides clinicians with (1) AFib in One Page, a 1-page document that provides guideline-based recommendations for rate/rhythm control and stroke prevention management; (2) Comparison of Anticoagulants, a 1-page document comparing the effectiveness and safety of the available anticoagulants for stroke prevention; and (3) Anticoagulation Dosing Table, a 1-page document providing dosage recommendations for all of the available anticoagulants. Additionally, some participants were unclear on how to rate their patients' AF using the Severity of AF (SAF) scale found on the initial assessment and follow-up visit forms. In response, a "learn more" button was added that provided definitions for each SAF class.

Lastly, participants described issues relating to the functionality of the EMR toolkit. For example, several links and buttons embedded in the toolkit did not work properly. In response, these functionality-related issues were fixed by our programmer.

Phase 2 Usability Testing: Key Themes

The key themes identified from Phase 2 usability testing can be grouped into the following sections: features that would promote use of the EMR toolkit of AF, areas for improvement, and perceived impact of the toolkit on patient care and workflow. No major usability-related issues were identified in Phase 2 of usability testing.

Features That Would Promote Use of the EMR Toolkit for AF

Participants described several benefits to using the EMR toolkit for AF. Three main subthemes were identified: (1) ease of accessibility to AF-related information and resources is beneficial, (2) structured guide for AF management serves as a reminder, and (3) structure and format of the toolkit were easy to use and follow.

Ease of Accessibility to AF-Related Information and Resources is Beneficial

A common advantage of the EMR toolkit for AF described by users was the accessibility of AF-related information and resources provided to clinicians. Participants described that they liked the fact that information was available to them at the "touch of a button," as information pertaining to AF management could be conveniently found in the Patient Resources, Provider Resources, and Stroke & Bleeding Risk Calculator sections of the toolkit. A family physician described, "there's so much information and I love the provider resources as well because sometimes you think you're not sure about something." With respect to the Stroke & Bleeding Risk Calculator, one participant stated that she liked that "you can pull up the CHADS score easily so you don't have to remember the whole thing if you're not sure." Another family physician described the time-saving benefit of using the toolkit and the lengthy process of accessing information without the toolkit:

> *I think it would save me time, it saves time to have it embedded. I mean we certainly do have patient handouts in the EMR but I'd have to come here and go (to) handouts and then look for atrial [fibrillation]...it would be I assume scanned in under atrial fibrillation and then click it and then view it and then print it.*

From survey results, 91% (10/11) of participants believed the EMR toolkit for AF improved their ability to access the information they needed to provide AF care.

Structured Guide for AF Management Serves as a Reminder

Most participants felt that the structured guide for AF management served as a reminder for what to ask patients regarding their AF care. A family physician stated that the toolkit was "good because...like everything else it gives you a list [so] you don't forget what you're supposed to be check[ing] which especially on a busy day you tend to rush and miss stuff." One family physician described how the toolkit served as a cheat sheet and ensured a systematic approach was taken towards the management of patients with AF:

> *I think it's great, it is to me a real cheat sheet, like it makes sure you don't miss anything and that you do go through an organized, systematic approach to dealing with atrial [fibrillation]...I just think it's really efficient and it makes sure you...do what you're supposed to do.*

Structure and Format of the EMR Toolkit for AF Was Easy to Use and Follow

Participants felt that the EMR toolkit for AF was easy to use and follow. They described how the toolkit was clear, intuitive, and straightforward. A family physician stated "sometimes [with] forms you can't find what you need to do," but the EMR toolkit was "fine...it was very easy to use." Another family physician described the intuitive nature of the initial assessment form:

> *It was actually pretty intuitive...In terms of obtaining a history of atrial fibrillation, so asking about symptoms, asking about risk factors and then examining the patient and coming up with their stroke risk and plan, that is very intuitive flow.*

From survey results, 100% (11/11) of participants thought that aspects of the EMR toolkit for AF were easy to use. Most (82%, 9/11) participants felt that the EMR toolkit for AF was compatible with their typical workflow.

Areas for Improvement

Participants suggested changes that could be made to improve the toolkit and its uptake by clinicians. They expressed (1) a desire for a prompt to redo stroke and bleeding risk assessments when needed and (2) a need for more education and awareness about the toolkit.

Desire for a Prompt to Redo Stroke and Bleeding Risk Assessments When Needed

All participants expressed that it would be helpful to have a reminder in the EMR system to prompt them to redo a stroke and bleeding risk assessment when certain CHADS-related patient characteristics (ie, age, new comorbidities) changed. A family physician described how receiving reminders would be helpful as long as it wasn't too frequent:

> *I think [it would be helpful]...if it did it at the age 65 and 75, whenever the brackets are, not every birthday...Or if there was somehow when...they get a new diagnosis of something if it could prompt you to think of it that would be really helpful...*

Need for More Education and Awareness About the Toolkit

Participants described the need to provide education and awareness about the toolkit and its functionalities to clinicians. A family physician suggested the need for training and hands-on practice with the toolkit to gain familiarity with it:

> *I think it does take some time to get familiar with using these so it's like practice. So I think it would be best if people...had training or something...just some examples of case studies or something...you*

> *have a new patient with AFib, this is what you would do and then you would go here because for me to see this for the very first time this form to fill out I think would take a lot of time.*

From survey results, 50% (5/10) of participants thought an orientation session was necessary to introduce the AF EMR toolkit. Most (64%, 7/11) participants believed an orientation session would increase usage of the toolkit.

Perceived Impact of the Toolkit on Patient Care and Workflow

Intention to Continue Using the EMR Toolkit for AF

All interviewed participants expressed their intent to continue using the toolkit to help guide AF management. A family physician acknowledged the benefits of certain aspects of the toolkit but emphasized her preference for the stroke and bleeding risk calculator:

> *I mean in all honesty I think if there was one thing that I for sure I'll use is the stroke and bleeding risk piece of it. The initial assessment, I think I would continue to use just [so] I take an AFib history in a very organized way, but if I'm super busy and I kind of forget that one might be the first to go but definitely the stroke and bleeding risk I would in terms of guiding managing I definitely would use that.*

From survey results, 82% (9/11) of surveyed clinicians intended to continue using aspects of the AF EMR toolkit in the future. Overall, 91% (10/11) of participants would recommend the EMR toolkit for AF to other clinicians. However, remembering to use the toolkit was suggested as the main barrier to its use.

Increased Thoroughness and Quality of AF-Related Care

Interviewed participants expressed that the toolkit prompted them to provide a more thorough assessment for patients with AF. A family physician described how the toolkit "prompted me to do things that I probably maybe wouldn't have done." This finding was supported by survey results with the majority (82%, 9/11) of participants agreeing that the toolkit increased the thoroughness of their assessments of patients as recommended by AF guidelines. Most (73%, 8/11) participants believed the toolkit increased the quality of AF-related care received by their patients.

Discussion

Principal Findings

The EMR toolkit for AF was designed as a resource for primary care clinicians to help guide decision making and improve the proportion of patients who are receiving guideline-concordant AF care. Through our interdisciplinary and human factors approach, we were able to develop a toolkit that incorporated perspectives from individuals with different areas of expertise—family medicine, pharmacy, cardiology, and nursing—and that was optimized for use in a large, busy family practice environment.

As a result of usability testing, we were able to develop a user-centered intervention that met the needs of primary care clinicians by identifying potential problems and areas for improvement early in the development stages. Nine iterations of the toolkit were created in response to feedback from clinicians who participated in Phase 1 of the study to make the toolkit more effective; interface-related changes were made, additional AF-related resources were added, and functionality issues were fixed. After these improvements to the toolkit were made, clinicians provided positive feedback regarding the toolkit and its perceived impact. Clinicians expressed that the toolkit improved accessibility to AF-related information and resources, served as a reminder for guideline-concordant AF management, and was easy to use. Most clinicians intended to continue using the toolkit for patient care, which may be attributable to its user-centered design. With respect to impact on care, clinicians believed the toolkit increased the thoroughness of their assessments for patients with AF and improved the quality of AF-related care received by their patients.

The results of this pilot study informed the refinement of the toolkit to make it a more effective, holistic, and user-centered intervention. The final version of the toolkit will be implemented across primary care clinics in Ontario in a cluster-randomized controlled trial examining its impact on guideline-concordant AF care. Through this mixed-methods study, we were able to demonstrate the utility of a human factors approach to piloting and refining an intervention prior to wide-scale implementation.

Comparison With Prior Work

The importance of incorporating the end users' needs and workflow into the design of information technology (IT) interventions has been well documented [16-20]. Additionally, research has found that the usability of a system is a major theme impacting use of IT interventions [21,22]. Incorporating these principles, participants expressed positive feedback about the EMR toolkit for AF and an intention to continuing using the toolkit for AF management. The technology acceptance model also suggests that the perceived usefulness and ease-of-use of a technology influences end users' intentions to use the technology. Our study supports this theory. This study provides novel information on the utility of a human factors approach to the development of an IT intervention for AF in the primary care setting.

Limitations

This study has several limitations. The study was conducted in urban-based, academic family health centers, the sample size for the survey was small, and the selection of clinicians may have been biased toward those who are accepting of novel quality improvement interventions. As such, the results may not be generalizable to all primary care practices in Ontario. The study also sought out self-reported experiences, and as a result, participant responses may be impacted by response bias and recall bias. However, data were collected through three methods—interviews, observations, and surveys—which provided similar results, reinforcing the themes identified.

Conclusion

Although an experienced multidisciplinary team carefully designed the EMR toolkit, we found that the use of a human factors approach enabled the development of an intervention

that better met the needs of primary care clinicians. The positive feedback surrounding the EMR toolkit for AF and its perceived impact on patient care can be attributed to the adoption of a user-centered design that merged clinically important information about AF management with user needs. This study demonstrates the utility of a human factors approach to piloting and refining an intervention prior to wide-scale implementation.

Acknowledgments

NMI is supported by New Investigator Awards from the Canadian Institutes of Health Research and the Department of Family and Community Medicine at the University of Toronto.

Conflicts of Interest

None declared.

Multimedia Appendix 1

EMR toolkit screenshot.

[PDF File (Adobe PDF File), 233KB - humanfactors_v3i1e7_app1.pdf]

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: The Framingham Study. Stroke 1991 Aug;22(8):983-988 [FREE Full text] [Medline: 1866765]
2. Heart and Stroke Foundation. 2014 Sep. Atrial Fibrillation - Be Pulse Aware URL: http://www.heartandstroke.com/site/c.ikIQLcMWJtE/b.5052135/k.2C86/Heart_disease__Atrial_fibrillation.htm [accessed 2015-12-14] [WebCite Cache ID 6dmDHwlm5]
3. Gladstone DJ, Bui E, Fang J, Laupacis A, Lindsay MP, Tu JV, et al. Potentially preventable strokes in high-risk patients with atrial fibrillation who are not adequately anticoagulated. Stroke 2009 Jan;40(1):235-240 [FREE Full text] [doi: 10.1161/STROKEAHA.108.516344] [Medline: 18757287]
4. Skanes AC, Healey JS, Cairns JA, Dorian P, Gillis AM, McMurtry MS, Canadian Cardiovascular Society Atrial Fibrillation Guidelines Committee. Focused 2012 update of the Canadian Cardiovascular Society atrial fibrillation guidelines: Recommendations for stroke prevention and rate/rhythm control. Can J Cardiol 2012;28(2):125-136. [doi: 10.1016/j.cjca.2012.01.021] [Medline: 22433576]
5. Louis KM, Martineau J, Rodrigues I, Fournier M, Berbiche D, Blais N, et al. Primary care practices and determinants of optimal anticoagulation management in a collaborative care model. Am Heart J 2010 Feb;159(2):183-189. [doi: 10.1016/j.ahj.2009.11.015] [Medline: 20152215]
6. Peterson GM, Boom K, Jackson SL, Vial JH. Doctors' beliefs on the use of antithrombotic therapy in atrial fibrillation: Identifying barriers to stroke prevention. Intern Med J 2002;32(1-2):15-23. [Medline: 11783668]
7. Bungard TJ, Ghali WA, McAlister FA, Buchan AM, Cave AJ, Hamilton PG, et al. Physicians' perceptions of the benefits and risks of warfarin for patients with nonvalvular atrial fibrillation. CMAJ 2001 Aug 7;165(3):301-302 [FREE Full text] [Medline: 11517646]
8. Beyth RJ, Antani MR, Covinsky KE, Miller DG, Chren MM, Quinn LM, et al. Why isn't warfarin prescribed to patients with nonrheumatic atrial fibrillation? J Gen Intern Med 1996 Dec;11(12):721-728. [Medline: 9016418]
9. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: Observational study. BMJ 2001 Nov 24;323(7323):1218-1222 [FREE Full text] [Medline: 11719412]
10. Protheroe J, Fahey T, Montgomery AA, Peters TJ. The impact of patients' preferences on the treatment of atrial fibrillation: Observational study of patient based decision analysis. BMJ 2000 May 20;320(7246):1380-1384 [FREE Full text] [Medline: 10818030]
11. Bungard TJ, Ghali WA, McAlister FA, Buchan AM, Cave AJ, Hamilton PG, et al. The relative importance of barriers to the prescription of warfarin for nonvalvular atrial fibrillation. Can J Cardiol 2003 Mar 15;19(3):280-284. [Medline: 12677283]
12. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]
13. Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health Aff (Millwood) 2005;24(5):1103-1117 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1103] [Medline: 16162551]
14. Adams WG, Mann AM, Bauchner H. Use of an electronic medical record improves the quality of urban pediatric primary care. Pediatrics 2003 Mar;111(3):626-632. [Medline: 12612247]

15. Kohn L, Corrigan J, Donaldson M. Institute of Medicine. 1999. To Err Is Human - Building a Safer Health System URL: http://iom.edu/~/media/Files/Report%20Files/1999/To-Err-is-Human/To%20Err%20is%20Human%201999%20%20report%20brief.pdf [accessed 2015-01-14] [WebCite Cache ID 6Va4jXNx0]
16. Kjeldskov J, Skov MB, Stage J. A longitudinal study of usability in health care: Does time heal? Int J Med Inform 2010 Jun;79(6):e135-e143. [doi: 10.1016/j.ijmedinf.2008.07.008] [Medline: 18757234]
17. Lowry S, Ramaiah M, Patterson E, Brick D, Gurses A, Ozok A, et al. National Institute of Standards and Technology, US Department of Commerce. 2014. Integrating Electronic Health Records into Clinical Workflow: An Application of Human Factors Modelling Methods to Ambulatory Care URL: http://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7988.pdf [accessed 2015-01-14] [WebCite Cache ID 6Va5Emq33]
18. Karsh B. Beyond usability: Designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
19. 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]
20. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005 Feb;38(1):75-87 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.005] [Medline: 15694887]
21. 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]
22. Smelcer J, Miller-Jacobs H, Kantrovich L. Usability of electronic medical records. J Usability Stud 2009;4(2):70-84.

Abbreviations

AF: atrial fibrillation
EMR: electronic medical record

Edited by G Eysenbach; submitted 26.01.15; peer-reviewed by M Janda, M Johnston; comments to author 05.08.15; revised version received 19.08.15; accepted 07.10.15; published 17.02.16.

Please cite as:
Tran K, Leblanc K, Valentinis A, Kavanagh D, Zahr N, Ivers NM
Evaluating the Usability and Perceived Impact of an Electronic Medical Record Toolkit for Atrial Fibrillation Management in Primary Care: A Mixed-Methods Study Incorporating Human Factors Design
JMIR Human Factors 2016;3(1):e7
URL: http://humanfactors.jmir.org/2016/1/e7/
doi:10.2196/humanfactors.4289
PMID:27026394

Original Paper

How Regrouping Alerts in Computerized Physician Order Entry Layout Influences Physicians' Prescription Behavior: Results of a Crossover Randomized Trial

Rolf Wipfli[1*], PhD; Frederic Ehrler[1*], PhD; Georges Bediang[2], MD; Mireille Bétrancourt[3], PhD; Christian Lovis[1], MD, MPH

[1]Division of Medical Information Sciences, Department of Radiology and Medical Informatics, University Hospitals of Geneva, Geneva, Switzerland
[2]Division of ehealth and Telemedicine, Department of Radiology and Medical Informatics, University Hospitals of Geneva, Geneva, Switzerland
[3]School of Psychology and Education, Faculty of Psychology and Education Sciences, University of Geneva, Geneva, Switzerland
[*]these authors contributed equally

Corresponding Author:
Frederic Ehrler, PhD
Division of Medical Information Sciences
Department of Radiology and Medical Informatics
University Hospitals of Geneva
Rue Gabrielle-Perret-Gentil 4
Geneva,
Switzerland
Phone: 41 22 372 8697
Fax: 41 22 372 6255
Email: frederic.ehrler@hcuge.ch

Abstract

Background: As demonstrated in several publications, low positive predictive value alerts in computerized physician order entry (CPOE) induce fatigue and may interrupt physicians unnecessarily during prescription of medication. Although it is difficult to increase the consideration of medical alerts by physician through an improvement of their predictive value, another approach consists to act on the way they are presented. The interruption management model inspired us to propose an alternative alert display strategy of regrouping the alerts in the screen layout, as a possible solution for reducing the interruption in physicians' workflow.

Objective: In this study, we compared 2 CPOE designs based on a particular alert presentation strategy: one design involved regrouping the alerts in a single place on the screen, and in the other, the alerts were located next to the triggering information. Our objective was to evaluate experimentally whether the new design led to fewer interruptions in workflow and if it affected alert handling.

Methods: The 2 CPOE designs were compared in a controlled crossover randomized trial. All interactions with the system and eye movements were stored for quantitative analysis.

Results: The study involved a group of 22 users consisting of physicians and medical students who solved medical scenarios containing prescription tasks. Scenario completion time was shorter when the alerts were regrouped (mean 117.29 seconds, SD 36.68) than when disseminated on the screen (mean 145.58 seconds, SD 75.07; *P*=.045). Eye tracking revealed that physicians fixated longer on alerts in the classic design (mean 119.71 seconds, SD 76.77) than in the centralized alert design (mean 70.58 seconds, SD 33.53; *P*=.001). Visual switches between prescription and alert areas, indicating interruption, were reduced with centralized alerts (mean 41.29, SD 21.26) compared with the classic design (mean 57.81, SD 35.97; *P*=.04). Prescription behavior (ie, prescription changes after alerting), however, did not change significantly between the 2 strategies of display. The After-Scenario Questionnaire (ASQ) that was filled out after each scenario showed that overall satisfaction was significantly rated lower when alerts were regrouped (mean 4.37, SD 1.23) than when displayed next to the triggering information (mean 5.32, SD 0.94; *P*=.02).

Conclusions: Centralization of alerts in a table might be a way to motivate physicians to manage alerts more actively, in a meaningful way, rather than just being interrupted by them. Our study could not provide clear recommendations yet, but provides objective data through a cognitive psychological approach. Future tests should work on standardized scenarios that would enable

to not only measure physicians' behavior (visual fixations and handling of alerts) but also validate those actions using clinical criteria.

(JMIR Human Factors 2016;3(1):e15) doi:10.2196/humanfactors.5320

KEYWORDS

medical order entry systems; clinical decision support systems; adverse drug reaction reporting systems; User-Computer Interface; eye tracking

Introduction

Background

Clinical information systems offer integrated views on patients' medical condition aiming at facilitating diagnosis. In order to facilitate decision making, systems increasingly not only communicate factual information but also interactively support clinical decision process. Clinical decision support systems are most often used in computerized physician order entry (CPOE) systems and typically include alerts for drug interactions. In such systems, medical alerts warn physicians when a prescription leads to a potential harmful situation for the patient's health. Drug-related alerts can be classified into two main categories: (1) basic alerts, which verify that dosage, route of administration, and frequency of prescriptions are within the recommended range, and (2) advanced alerts, which rely on information from patients' electronic medical records to provide personalized advice [1].

A positive effect of medical alerts on prescription behavior and, to a smaller extent, on patient outcomes can be found in the literature [2, 3]. However, research has shown that these alerts are still underutilized despite their great potential. According to a meta-study, more than half of these alerts are overridden [4]. Possible explanations for the low compliance include the lack of specificity of the alerts, their poor inclusion in the clinical workflow, and usability issues.

The human-computer interface is identified as a determining factor for improving clinical information systems (CISs) [5]. In current vendor CISs, alerts and reminders are typically displayed as pop-ups that interrupt the workflow or are displayed in the medical record using symbols to attract attention. Several guidelines [6, 7], often based on expert consensus, recommend what information has to be displayed and how this information should be conveyed. These guidelines advocate prioritization using different symbols and colors to reduce the number of alerts requiring acknowledgment and to display alerts spatially and temporally close to the triggering information. Another approach for improving usability is to reengineer the way alerts are presented and experimentally test with prototypes [8]. For example, the effect of interruptive and non-interruptive alerts on prescription behavior has been studied [9, 10].

Different approaches have been used to improve the effectiveness of medical alerts. Contextualizing the alerts [11] and eliminating those with low clinical evidence [12] or low severity [13] help to achieve a better specificity. Unfortunately, there is little consensus on what alerts are superfluous and can be removed from a system [14, 15]. Future research should aim at improving alerts' sensitivity and specificity, better adaptation of alerts to prescribers' personal needs, and reducing the number of alerts [16].

In our prior research [17], we learned about physicians' use of CPOE and alerts through a work analysis of the prescription activity at the University Hospitals of Geneva. The insight obtained during interviews revealed that physicians consult medical alerts only in rare, unfamiliar medical situations, ignoring them for numerous routine prescriptions. The study demonstrated that alert handling is an active process where physicians rely on the alerting system for only complex unfamiliar medical prescriptions. This made us realize that the alert handling and the prescription of medication can be considered as two different tasks, with the former likely to unnecessarily disturb the latter.

These observations led us to propose an alternative alert presentation layout inspired by the interruption management model [18]. This model describes how interruption stimuli such as medical alerts are processed by physicians. The model shows that physicians experience cognitive load when alerts are displayed, even when they are not handled. On the basis of these ideas, we proposed a new principle advising that active alerts should be displayed regrouped in a centralized area in the prescription layout where physicians can consult and manage them. Instead of an immediate interruption, we propose a negotiated interruption where physicians are informed of alerts but can choose when to handle them.

Study Objective

This study aimed to investigate whether centralizing alerts in a CPOE interface can lead to a reduction in the interruption of the prescription workflow without reducing the consideration of alerts by physicians.

Methods

Study Design

In order to compare 2 alert display strategies, 2 CPOE designs based on these principles have been compared in a crossover randomized controlled trial. In the first display design, alerts are displayed on the screen spatially proximate to the triggering information. In the second design, alerts are displayed centralized in one table. An eye-tracking device was employed for measuring inspection time on alerts and switches between alerts and prescription areas. Finally, the satisfaction questionnaire ASQ was used to measure user satisfaction with the 2 alert display strategies.

Scenarios and Alerts

Eight scenarios aiming to solicit medical reasoning were created. The scenario contained information about fictive patient identities (name, age, and sex) and instructions for prescription of medications. Each scenario provided 2 types of alerts. Some alerts were activated and visible from the beginning. Others were triggered depending on prescriptions when following the instructions. Alerts could have 3 levels of severity (increasing from 0 to 2) and 3 levels of urgency (increasing from 0 to 2). Alerts of severity level 0 are informative alerts such as "There is already a drug of the same therapeutic class." These alerts are relatively frequent in many situations, such as treatment of hypertension, are considered to be of "low importance," and are thus often ignored by the physicians. Level 1 alerts are considered as severe and must usually be taken into account. However, there might be several medical reasons to overcome the alerts. For instance, "The dosage of the drug is too high considering the renal function of the patient." Finally level 2 alerts such as "The patient is allergic (level anaphylactic) to this drug" are considered as very severe alerts, in the same group as severe interactions. These level 2 alerts should never be overpassed, except in very special situations requiring specific accesses. Severe and urgent alerts interrupted the workflow, whereas others were displayed on the screen without requiring any actions. The alerts were chosen in such a way that their different levels of severity, urgency, and modality were represented (see Multimedia Appendix 1).

Besides pharmacological alerts, physicians are confronted with other alerts and reminders during their use of the CIS. Systems warn about patients' allergies, increased hygiene measures when patient is infected, and even reminders to consult recent patient-related data provided by another hospital division, for example, about the availability of new laboratory results. We added such alerts to our prototype to represent the full range of alerts typical for a CIS.

Participants

Participants were recruited among physicians from the University Hospitals of Geneva and the Faculty of Medicine at the University of Geneva. Students were eligible to be included as participants when they have had some work experience as trainees in the hospital and have used the hospital's CPOE already. The faculty of psychology of the University of Geneva approved the ethical aspects of the study that was a part of a larger PhD thesis. Because the purpose of this study was to examine the effect only on the providers, trial registration was unnecessary.

Study Flow

Compared Computerized Physician Order Entry Designs

To test the hypotheses, 2 CPOE layouts have been designed based on the alerts display strategy and compared in a crossover study. The 2 designs are based on a common base of the hospital's CPOE system. In such a system (as seen in Figures 1 and 2), the column on the left allows physicians to choose the drug to prescribe. The physician can type the beginning of the name of the drug and is provided with a list of suggestions of drugs available at the hospitals. In the center of the screen, different options are available to define the dose and the frequency of administration of the drug as well as the beginning and ending dates of administration. A text area is reserved for additional comments.

Figure 1. Classic computerized physician order entry design.

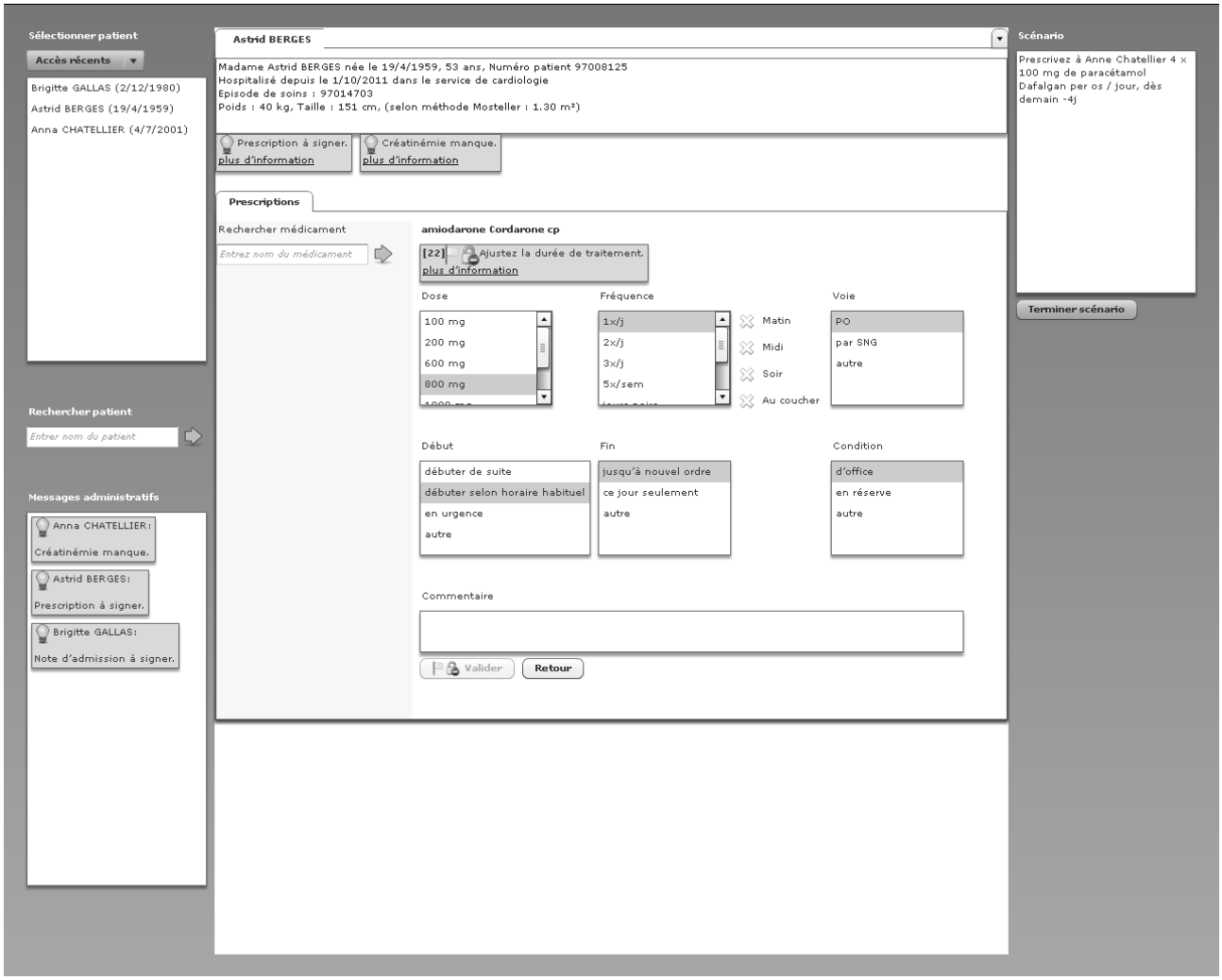
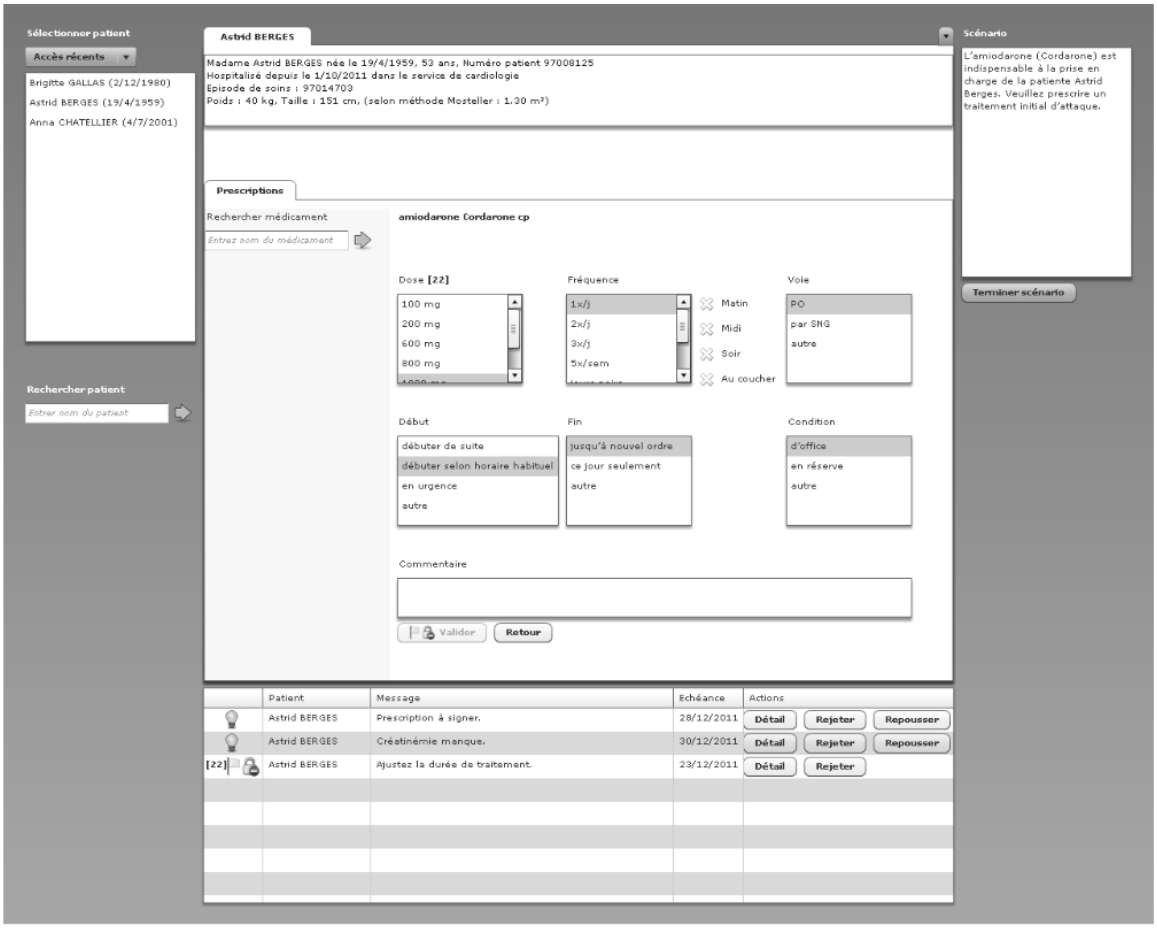


Figure 2. Cognitively engineered computerized physician order entry design.



Classic Design

In classic (CL) design (Figure 1), the *alert position* is integrated, which means it is located near the triggering information. Physicians can open an alert by clicking the link "more information." Once opened, a detailed view is displayed in a pop-up window (not visible on the figure but located at the center of the screen) in which physicians can reject or postpone the alert. General alerts are located on the left side. Administrative alerts and reminders are on the top. Prescription-specific alerts are close to the triggering alerts. Clicking on alerts would open a pop-up with alert information.

Cognitively Engineered Design

In cognitively engineered (CE) design (Figure 2), the *alerts are regrouped in a defined location*. All alerts are centralized in a table at the bottom of the screen where physicians can interact with them. Three options are available to the physicians. They can click the button "detail" to open the alert in a detailed view or they can reject or postpone the alert.

Randomization Strategy

There is a controlled variable named *scenario group*. In each scenario group (A and B), there are 4 scenarios describing a medical case. The factors *scenario group* and *type of presentation* are randomized block wise (see Figure 3). The order of the 4 scenarios within the scenario group was not randomized, which enabled us to create scenarios using the same fictive patient twice.

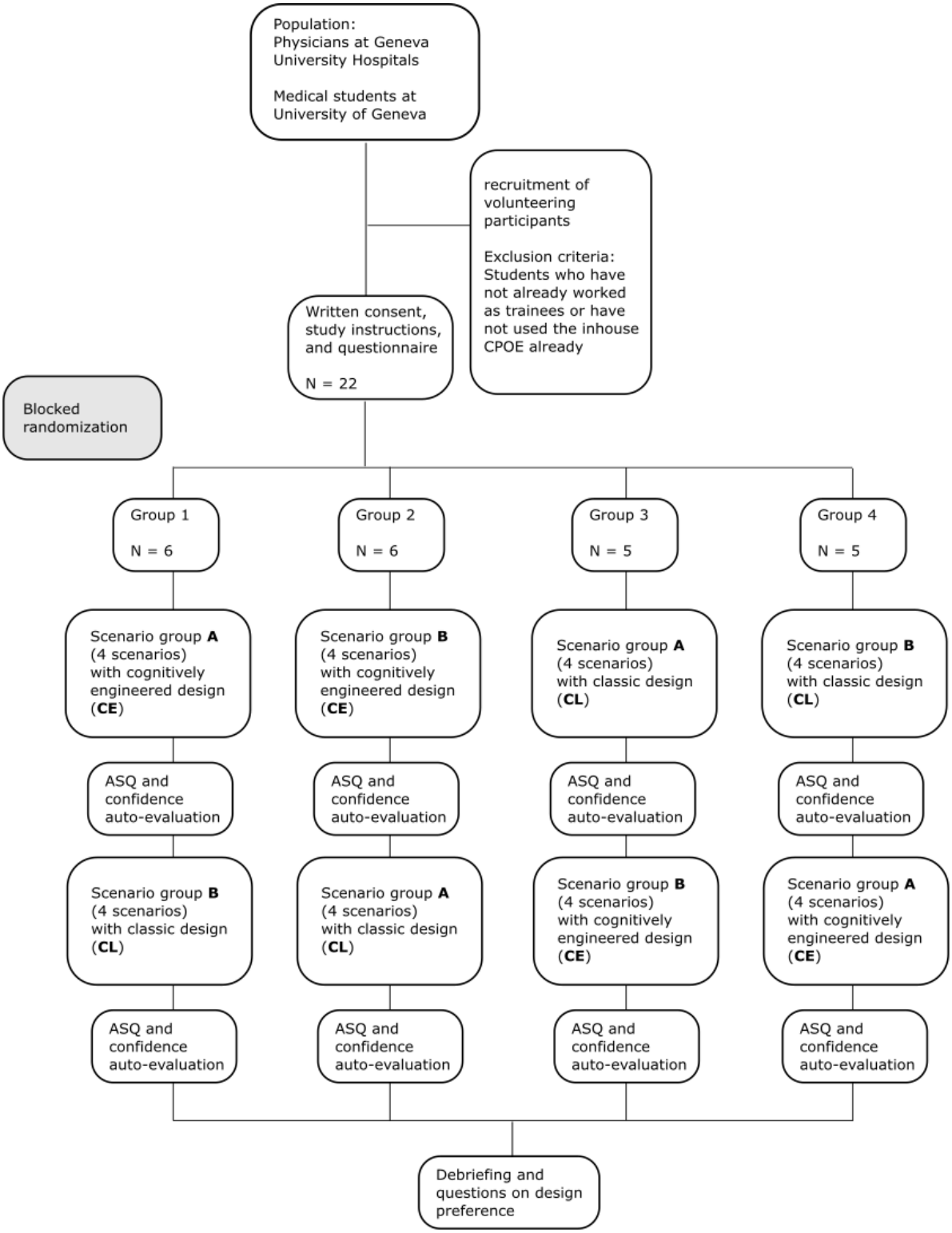
Each participant performed the test individually in a dedicated test room under the supervision of an experimenter. The monitor with which the participant interacted was connected to a Tobii T120 Eye Tracker. This eye tracker is an infrared corneal reflection–based device with a data rate of 120 Hz and the accuracy of 0.5 degrees. The screen size is 17 inches (43.18 cm) with a resolution of 1280×1024 pixels. The infrared emitters and the infrared camera are integrated in the monitor. The interaction with the eye tracker is similar to that with a standard work place computer.

In a first step, the participant was briefed about the study goal and filled out a consent form. Subsequently, he or she was interviewed on his or her previous use of electronic prescription systems. Afterward, the eye tracker was calibrated to the participant's eyes. Once the calibration was completed, the participant was instructed on how to use the interface. Subsequently, the online phase of the experiment began.

The participants started with either the CL design or the CE design. After completion of each scenario, the participants were asked by the system to what extent (in percentage) they thought they had accomplished the task. Then the participants repeated the procedure presented in this paragraph with the second design.

When the participants finished the online phase of the test, they were asked in a posttest interview about their general impression and their preference for one of the designs.

Figure 3. Study flow. CPOE: computerized physician order entry; ASQ: After-Scenario Questionnaire.



Outcome Measures

Scenario Completion Time

Our primary outcome measure was scenario completion time. Scenario completion time refers to the time difference in seconds between the start of the scenario to the time they proceed to the following scenario. Participants could proceed to the next scenario whenever they considered the current scenario to be completed.

Prescription and Alert Handling

All interactions with the prescription prototype were captured. Every time the participant clicked a button to open, postpone, or reject alerts or to validate or sign prescriptions, a corresponding log entry was created.

Visual Switches

We computed the frequency of the physician's visual focus switch between the prescription area and the different alert areas using an eye tracker. Only direct transitions from one area of interest to the other were counted. The alert areas were either the alert table in the CE design or the areas where the alerts would appear in the CL design.

Fixation Duration

The duration of fixations adds up participants' visual fixations within either alert areas or prescription areas on the screen. Moreover, in the CL design the fixation duration on the pop-ups is added to the total fixation duration.

Satisfaction and Confidence

A user satisfaction questionnaire, the ASQ [19], was applied after each type of design. The ASQ is a 3-item questionnaire with a 7-point Likert scale ranging from *strongly agree* to *strongly disagree*. After each scenario, participants could rate the degree of confidence in the correctness of the prescription they had performed. The question was "To what degree in percent do you think that you have accomplished the task?". Participants rated on a scale from 1 to 100 in steps of 10.

Data Analysis

Data analysis was conducted with R statistics version 3.1.2. Shapiro-Wilk test for testing distributions for normality was used. A significance level of .05 was used for analyses. When conditions for a parametric test were met, a 2-sided paired student *t*test was used. For nonparametric tests, a paired Wilcoxon signed rank test was used. The Likert scales for the ASQ questionnaire were considered to be continuous [20]. For design preference, we used a binominal test with the assumption of a theoretical number of 10.5 supporters in each group.

Results

Participants

The sample consisted of 22 medical students and physicians, among whom 7 were women. Three participants were medical students, 8 were novice physicians who had their medical diploma for less than 4 years, and 11 were experienced physicians who had their medical diploma for at least 4 years. The participants had a mean experience of 2 years and 9 months with electronic prescription systems.

The physicians work in the division of general internal medicine (12), service of eHealth (2), orthopedic surgery and trauma (1), pediatric orthopedics (1), palliative medicine (1), otolaryngology (1), and medical-economic analysis (1). Concerning the physicians' roles, there were 12 resident physicians, 3 attending physicians, 2 deputy heads of divisions, and 3 from the informatics division.

One participant had to be excluded from the analysis for perceptual and behavioral data. Only 40 % of eye-movement samples could be captured and no data were logged for interactions with the CL design for this participant. Data from the ASQ, however, were included in the analysis. The 21 remaining participants had an average of 81.86 % of eye movements recorded (SD 11.15). In Table 1, all following results are summarized.

Scenario Completion Time

The execution time with the CL design was on average 145.58 (SD 75.07) seconds per scenario. Using the CE design, the execution was shorter with 117.29 (SD 36.68) seconds. This difference was significant with P=.045.

Fixation Duration

The inferential analysis revealed a significant positive difference in the duration of fixations on medical alerts in the CL design (mean 119.71 seconds, SD 76.77) compared with the duration of fixations on medical alerts in the CE design (mean 70.58 seconds, SD 33.53; P=.001).

Visual Switches

The number of switches between any of the medical alerts in the CL design and the prescription area showed that there was an average of 57.81 (SD 35.97) switches per scenario. In the CE design there were 41.29 (SD 21.26) switches per scenario between the prescription area and the table containing the medical alerts, which is significantly lower than in the CL design (P=.04).

Influence of the Physician's Experience With the Cpoe

A Kendall tau test evaluated the influence of the variables "experience with CPOE" (in months) and "medical experience group" (medical student, novice physician, expert physician) on scenario completion time, fixation duration, and visual switches. In all 3 cases, there was no significant correlation.

Prescription and Alerts Handling

Because only 7 alerts were postponed and 15 participants never clicked on a postpone button, this variable is excluded from the analysis. Participants opened significantly more alerts in the CL design (mean 7.10, SD 4.25) than in the CE design (mean 4.35, SD 3.12; P=.001). Participants rejected significantly more alerts while using the CE design (mean 1.86, SD 1.39; P=.01) than the CL design (mean 0.67, SD 0.91).

Furthermore, we counted the number of times participants signed or validated prescriptions and added up their occurrences. The difference between the CL design (mean 8.67, SD 2.57) and the CE design (mean 7.95, SD 2.65) is not significant. The number of corrected prescriptions was counted; they include the accumulated numbers of removed pending prescriptions or stopped signed prescriptions as well as the number of times the participants removed all pending prescriptions or stopped all signed prescriptions. There was no significant difference between the CL design (mean 2.90, SD 2.17) and the CE design (mean 2.33, SD 1.96).

A test for correlations (Kendall's tau) with the factors "experience with CPOE" (in months) and "medical experience group" (medical student, novice physician, expert physician) showed no significant correlations with the factors "number of alerts opened in CL design," "number of alerts opened in CE design," "number of alerts rejected in the CL design," and "number of alerts rejected in CE design."

Table 1. Overview of all results.

Performance indicators	CL[a] design mean (SD)	CE[b] design mean (SD)	Test and significance	
			V_{df}: paired Wilcoxon signed rank test	*P*-value
Scenario completion time, seconds	145.58 (75.07)	117.29 (36.68)	V_{20}= 173	*P*=.045
Fixation duration on alerts, seconds	119.71 (76.77)	70.58 (33.53)	V_{20}= 204	*P*=.001
Number of switches, N	57.81 (35.97)	41.29 (21.26)	V_{20}= 175	*P*=.04
Alerts opened, N	7.10 (4.25)	4.35 (3.12)	V_{20}= 176	*P*=.001
Alerts rejected, N	0.67 (0.91)	1.86 (1.39)	V_{20}= 13	*P*=.01
Prescriptions validated or signed, N	8.67 (2.57)	7.95 (2.65)	t_{20}= 0.831	*P*=.42
Prescriptions corrected, N	2.90 (2.17)	2.33 (1.96)	V_{20}= 92	*P*=.47
ASQ[c] ease of use (1-7); 1 = worst, 7 = best	5.36 (1.14)	4.64 (1.53)	V_{21}= 125	*P*=.08
ASQ efficiency (1-7); 1 = worst, 7 = best	5.54 (1.10)	4.64 (1.43)	V_{21}= 129	*P*=.06
ASQ support (1-7); 1 = worst, 7 = best	5.04 (1.10)	3.86 (1.81)	V_{21}= 146	*P*=.04
ASQ total (1-7); 1 = worst, 7 = best	5.32 (0.94)	4.37 (1.23)	V_{21}= 183	*P*=.02
Confidence level (1-100); 1 = worst, 100 = best	71.97 (15.63)	66.44 (17.68)	t_{21}= 1.65	*P*=.11

[a]CL: classic.

[b]CE: cognitively engineered.

[c]ASQ: After-Scenario Questionnaire.

Satisfaction and Confidence

The factor “ease of use” was rated higher for the CL design (mean 5.36, SD 1.14) than for the CE design (mean 4.64, SD 1.53) but failed to reach a significant level (*P*=.08). The factor “efficiency” was rated higher for the CL design (mean 5.54, SD 1.10) than for the CE design (mean 4.64, SD 1.43; *P*=.06). This difference was not significant. The third factor “support” was rated significantly higher in the CL (mean 5.45, SD 1.10) compared with the CE design (mean 3.86, SD 1.81; *P*=.04). Finally, the “overall satisfaction” was significantly rated higher (*P*=.02) for the CL design with an average of 5.32 (SD 0.94) compared with an average of 4.37 (SD 1.23) for the CE design.

During the posttest interviews, 13 participants said they preferred the CL design and 8 said they preferred the CE design. The binomial test revealed that there was no significant difference.

Furthermore, a questionnaire evaluated whether the participants felt more confident in the solution they choose for the scenarios in either the CL or CE design. The participants had a slightly higher confidence in their solutions in the CL design (mean 71.97, SD 15.63) than in the CE design (mean 66.44, SD 17.68). The difference was not significant.

Discussion

Our new interface design was built on the assumption that physicians get less distracted by the CE design compared with the CL design. Participants switched significantly more often between the primary task (drug prescription) and the secondary task (alert handling) in the CL design. This reduction in the number of interruptions was accompanied by significantly shorter scenario completion time.

There is no definitive answer whether the reduced attention on the alerts, measured by alert fixation duration, and the fewer times participants opened alerts in the CE design are advantageous or disadvantageous. On the one hand, it could be argued that alerts in the CE design are not seen and therefore not opened. Consequently, they do not fulfill their function of warning the physician. On the other hand, there is evidence from a prior ethnographic study [17] that physicians are not driven by these alerting systems but rather consult them in case of uncertain conditions. In this latter case, it could be argued that alerts should not divert attention more than necessary from the prescription task. This second assumption is also supported by the fact that a similar number of corrective actions have been found in the 2 designs. Therefore, even if the participants clicked on more alerts and focused their attention more often on alerts in the CL design, it had no effect on how they responded to the alerts.

Significantly fewer alerts were rejected in the CL design than in the CE design. This is probably because participants could reject the alerts directly in the CE design without opening any alert.

It could seem surprising that participants considered the CL design to be more efficient because the results proved that they were more efficient with the CE design. However, this difference between perceived time and actual time is not new [21]. Overall, the participants were more satisfied with the CL design.

An important limitation of this study is the strong similarity between the CL design and the current CIS at the University Hospitals of Geneva. Thus, physicians were used to the CL design, which might influence satisfaction ratings, alert fixation, and handling. This fact does not prevent the generalization of our findings. The cognitively engineered design presented in this study can be applied to other CPOE systems and might advocate the use of centralized alerts.

We did not evaluate whether participants looked at alerts more frequently or handled them more frequently depending on their urgency or severity. A future test could examine in detail the different types of alerts. Moreover, tests should work on standardized scenarios that would enable us to not only measure physicians' actions (visual fixations and handling of alerts) but also validate those actions regarding clinical criteria. The alerts used in this study are conceived to be representative in their type, not in their frequency. For this reason, comparisons of our results with results from other studies reporting alert acceptance rates in a real clinical environment are not valid. Still, this study shows that the centralization of alerts influences workflow interruptions. It contradicts the proximity principle [6], which states that alerts should be close to the triggering text, but increases physicians' readiness to reject alerts that are irrelevant in their point of view.

In this work, we rely on theoretical knowledge on decision making and cognitive load to develop a new user interface for CPOE. Our formal measures, based on eye tracking, could demonstrate that following some simple design principles can affect alert handling. Centralizing alerts and making it possible to handle them in an active way reduce physicians' workflow interruptions without modifying their prescription decisions. We did not measure the quality of the medical decision making, which should be done in a future study. When the quality standards can be met, such design principles, based on scientific measures, can be used to improve the prescription behavior and, in a future step, patient safety.

Authors' Contributions

RW and FW have contributed equally. The study was initiated by CL and MB. The experimental study was designed by RW and supervised by CL and MB. RW conducted the experiment and performed data analysis. GB designed the medical scenarios. The manuscript was drafted by RW and reviewed and revised by FE, GB, CL, and MB. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Scenarios and alerts.

[PDF File (Adobe PDF File), 34KB - humanfactors_v3i1e15_app1.pdf]

References

1. Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, et al. Medication-related clinical decision support in computerized provider order entry systems: a review. J Am Med Inform Assoc 2007;14(1):29-40 [FREE Full text] [doi: 10.1197/jamia.M2170] [Medline: 17068355]
2. Schedlbauer A, Prasad V, Mulvaney C, Phansalkar S, Stanton W, Bates DW, et al. What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? J Am Med Inform Assoc 2009;16(4):531-538 [FREE Full text] [doi: 10.1197/jamia.M2910] [Medline: 19390110]
3. Garg AX, Adhikari Neill K J, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
4. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
5. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
6. Phansalkar S, Edworthy J, Hellier E, Seger DL, Schedlbauer A, Avery AJ, et al. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. J Am Med Inform Assoc 2010;17(5):493-501 [FREE Full text] [doi: 10.1136/jamia.2010.005264] [Medline: 20819851]
7. Payne TH, 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]
8. Cornu P, Steurbaut S, Gentens K, Van de Velde R, Dupont AG. Pilot evaluation of an optimized context-specific drug-drug interaction alerting system: A controlled pre-post study. Int J Med Inform 2015 Sep;84(9):617-629. [doi: 10.1016/j.ijmedinf.2015.05.005] [Medline: 26051616]

9. Hsieh TC, Kuperman GJ, Jaggi T, Hojnowski-Diaz P, Fiskio J, Williams DH, et al. Characteristics and consequences of drug allergy alert overrides in a computerized physician order entry system. J Am Med Inform Assoc 2004;11(6):482-491 [FREE Full text] [doi: 10.1197/jamia.M1556] [Medline: 15298998]
10. Scott GP, Shah P, Wyatt JC, Makubate B, Cross FW. Making electronic prescribing alerts more effective: scenario-based experimental study in junior doctors. J Am Med Inform Assoc 2011;18(6):789-798 [FREE Full text] [doi: 10.1136/amiajnl-2011-000199] [Medline: 21836158]
11. Jung M, Riedmann D, Hackl WO, Hoerbst A, Jaspers MW, Ferret L, et al. Physicians' perceptions on the usefulness of contextual information for prioritizing and presenting alerts in Computerized Physician Order Entry systems. BMC Med Inform Decis Mak 2012;12:111 [FREE Full text] [doi: 10.1186/1472-6947-12-111] [Medline: 23031275]
12. Luna D, Otero V, Canosa D, Montenegro S, Otero P, de Quirós FG. Analysis and redesign of a knowledge database for a drug-drug interactions alert system. Stud Health Technol Inform 2007;129(Pt 2):885-889. [Medline: 17911843]
13. Paterno MD, Maviglia SM, Gorman PN, Seger DL, Yoshida E, Seger AC, et al. Tiering drug-drug interaction alerts by severity increases compliance rates. J Am Med Inform Assoc 2009;16(1):40-46 [FREE Full text] [doi: 10.1197/jamia.M2808] [Medline: 18952941]
14. van der Sijs H, Aarts J, van Gelder T, Berg M, Vulto A. Turning off frequently overridden drug alerts: limited opportunities for doing it safely. J Am Med Inform Assoc 2008;15(4):439-448 [FREE Full text] [doi: 10.1197/jamia.M2311] [Medline: 18436915]
15. Missiakos O, Baysari MT, Day RO. Identifying effective computerized strategies to prevent drug-drug interactions in hospital: A user-centered approach. Int J Med Inform 2015 Aug;84(8):595-600. [doi: 10.1016/j.ijmedinf.2015.04.001] [Medline: 25912258]
16. Coleman JJ, van der Sijs H, Haefeli WE, Slight SP, McDowell SE, Seidling HM, et al. On the alert: future priorities for alerts in clinical decision support for computerized physician order entry identified from a European workshop. BMC Med Inform Decis Mak 2013;13:111 [FREE Full text] [doi: 10.1186/1472-6947-13-111] [Medline: 24083548]
17. Wipfli R, Betrancourt M, Guardia A, Lovis C. A qualitative analysis of prescription activity and alert usage in a computerized physician order entry system. Stud Health Technol Inform 2011;169:940-944. [Medline: 21893884]
18. McFarlane D, Latorella K. The Scope and Importance of Human Interruption in Human-Computer Interaction Design. Human-Comp. Interaction 2002 Mar 1;17(1):1-61. [doi: 10.1207/S15327051HCI1701_1]
19. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
20. Carifio J, Perla R. Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. J. of Social Sciences 2007 Mar 1;3(3):106-116. [doi: 10.3844/jssp.2007.106.116]
21. Fraisse P. Perception and estimation of time. Annu Rev Psychol 1984;35:1-36. [doi: 10.1146/annurev.ps.35.020184.000245] [Medline: 6367623]

Abbreviations

ASQ: After-Scenario Questionnaire
CE: cognitively engineered
CIS: clinical information system
CL: classic
CPOE: computerized physician order entry

Edited by G Eysenbach; submitted 04.11.15; peer-reviewed by J Aarts, J Adjedj, N Griffon, L Griebel; comments to author 23.01.16; revised version received 13.03.16; accepted 30.03.16; published 02.06.16.

Please cite as:
Wipfli R, Ehrler F, Bediang G, Bétrancourt M, Lovis C
How Regrouping Alerts in Computerized Physician Order Entry Layout Influences Physicians' Prescription Behavior: Results of a Crossover Randomized Trial
JMIR Human Factors 2016;3(1):e15
URL: http://humanfactors.jmir.org/2016/1/e15/
doi:10.2196/humanfactors.5320
PMID:27255612

Original Paper

Usability Testing and Adaptation of the Pediatric Cardiovascular Risk Reduction Clinical Decision Support Tool

Pamela A Williams[1], PhD, MS; Robert D Furberg[1], PhD, MBA; Jacqueline E Bagwell[1], MS, MMCi; Kenneth A LaBresh[1], MD

RTI International, Social Policy, Health and Economics Research Unit, Research Triangle Park, NC, United States

Corresponding Author:
Pamela A Williams, PhD, MS
RTI International
Faculty of Health
Social Policy, Health and Economics Research Unit
3040 East Cornwallis Road
Research Triangle Park
North Carolina, 27709-2194
United States
Phone: 1 919-316-3936
Fax: 1 919-541-8801
Email: pamwilliams@rti.org

Abstract

Background: Cardiovascular disease (CVD) is 1 of the leading causes of death, years of life lost, and disability-adjusted years of life lost worldwide. CVD prevention for children and teens is needed, as CVD risk factors and behaviors beginning in youth contribute to CVD development. In 2012, the National Heart, Lung, and Blood Institute released their "Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents" for clinicians, describing CVD risk factors they should address with patients at primary care preventative visits. However, uptake of new guidelines is slow. Clinical decision support (CDS) tools can improve guideline uptake. In this paper, we describe our process of testing and adapting a CDS tool to help clinicians evaluate patient risk, recommend behaviors to prevent development of risk, and complete complex calculations to determine appropriate interventions as recommended by the guidelines, using a user-centered design approach.

Objective: The objective of the study was to assess the usability of a pediatric CVD risk factor tool by clinicians.

Methods: The tool was tested using one-on-one in-person testing and a "think aloud" approach with 5 clinicians and by using the tool in clinical practice along with formal usability metrics with 14 pediatricians. Thematic analysis of the data from the in-person testing and clinical practice testing identified suggestions for change in 3 major areas: user experience, content refinement, and technical deployment. Descriptive statistical techniques were employed to summarize users' overall experience with the tool.

Results: Data from testers showed that general reactions toward the CDS tool were positive. Clinical practice testers suggested revisions to make the application more user-friendly, especially for clinicians using the application on the iPhone, and called for refining recommendations to be more succinct and better tailored to the patient. Tester feedback was incorporated into the design when feasible, including streamlining data entry during clinical visits, reducing the volume of results displayed, and highlighting critical results.

Conclusions: This study found support for the usability of our pediatric CVD risk factor tool. Insights shared about this tool may be applicable for designing other mHealth applications and CDS tools. The usability of decision support tools in clinical practice depends critically on receiving (ie, through an accessible device) and adapting the tool to meet the needs of clinicians in the practice setting.

(JMIR Hum Factors 2016;3(1):e17) doi:10.2196/humanfactors.5440

KEYWORDS

adaptation; cardiovascular diseases; clinical decision support; decision aids; guidelines; mHealth; pediatrics; risk factors; usability

Introduction

By 2020, cardiovascular disease (CVD) is projected to rank first in frequency among causes of death, years of life lost, and disability-adjusted years of life lost worldwide [1]. Because risk factors and behaviors that begin in youth can contribute to CVD later in life, prevention needs to start with children and teenagers [2]. Recognizing this need, in 2012 the National Heart, Lung, and Blood Institute (NHLBI) [3] released its 402-page, evidence-based "Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents." The comprehensive guidelines describe CVD risk factors that clinicians should address with patients from birth through 21 years of age, and with their families. The guidelines also include recommendations for clinicians on CVD risk factors, such as diet, physical activity, tobacco, blood pressure (BP), lipids and lipoproteins, and overweight and obesity, as well as the influence of family history on risk factor management. However, barriers related to lack of knowledge, effective systems, and support often delay uptake of new guidelines by clinicians [4]. Additionally, clinicians frequently have multiple prevention topics to discuss with patients [5], which may leave little time to add exploring CVD risk factors during primary care visits [6].

To overcome some of the barriers and to support clinicians in implementing the NHLBI CVD guidelines, we developed a comprehensive, multifaceted intervention that includes practice-based clinical champions, monthly collaborative webinars to support practice change, and a tool kit to support guideline implementation. The tool kit comprises a patient and family workbook to support patients in making behavior changes, guideline summary materials that include recommendations for clinicians, and a clinical decision support (CDS) tool [7-10]. CDS tools can improve clinician adherence to guidelines [11]. To assist clinicians in prioritizing the topics most important to individual patients to reduce their specific CVD risk, we developed the Pediatric Cardiovascular Risk Reduction CDS Tool, which provides a concise presentation of guideline information and tools to help clinicians perform the complex assessment of CVD risk factors within the clinical workflow. The tool allows clinicians to input patient and family history, determine guideline-specific recommendations for individual patients, and calculate both individual and combined body mass index (BMI) and BP percentiles. It also supports interpretation of specific laboratory results, including lipid screening values, for planning targeted follow-up visits [10].

The overarching aim of the CDS tool is to provide clinicians at the point of care with actionable, individually relevant recommendations for patients drawn from the comprehensive NHLBI CVD guidelines. We utilized a user-centered design approach in which users participated in pretesting and were involved with refinement of the design throughout the entire development process [12], including creating the content [8], designing and programming [7], pretesting (discussed in this paper), and conducting the experimental study [13]. Overall, we used a feature-driven development approach where primary components of the CDS tool were initially developed and tested independently, refined [7,8,14] and progressively integrated, and tested again.

After we created the content for the CDS tool, we developed other elements using key user-interface and user-experience design principles such as giving the user control, empowering the user, and allowing exploration and browsing; providing immediate feedback and the option of help at any point, and defining terminology used in the app; keeping the interface consistent, with active buttons in the same place throughout [12]; and incorporating actionable feedback related to the user experience and content refinement. We discuss these elements in this paper as we describe the process we used to pretest and adapt the CDS tool to help clinicians implement the guidelines. Research has shown the utility of conducting usability testing to adapt tools to maximize usability [15]. The aim of this study was to examine the usability of the Pediatric Cardiovascular Risk Reduction CDS Tool.

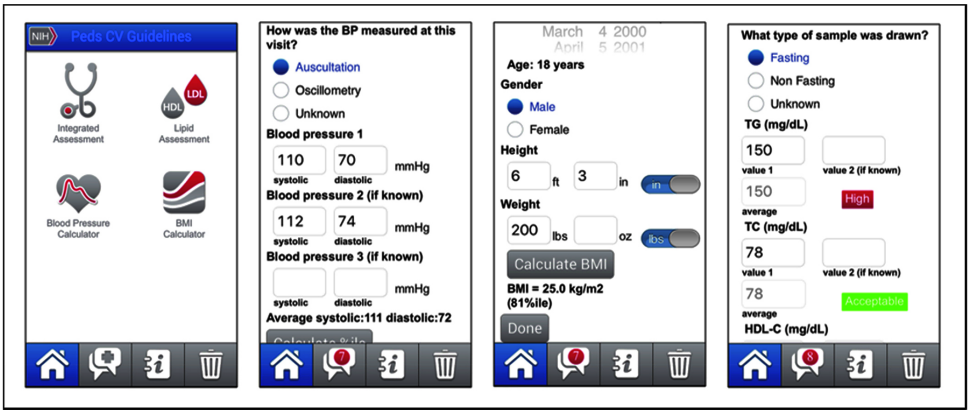
Methods

The Pediatric Cardiovascular Risk Reduction CDS Tool consists of a screening (integrated risk) assessment, BMI and BP calculators, and a lipid assessment instrument (Figure 1) and provides clinicians with a patient summary and NHLBI recommendations based on the patient's risk factor information input by the user. The integrated risk assessment module asks users to enter data to assess BMI (ie, patient's date of birth, gender, height, and weight), BP (ie, the patient's systolic and diastolic BP), and other risk factors (eg, whether the patient has dyslipidemia), and then provides users with a patient summary and NHLBI recommendations related to family history, nutrition and diet, physical activity, tobacco exposure, lipids, and overweight and obesity. It also provides supportive actions to take, based on the patient's risk factor information.

For the BMI and BP calculators and the lipid assessment instrument, if the user has already entered information in other modules (eg, integrated risk assessment), then the app will display the relevant, previously entered data in this module (eg, for the lipid assessment and BMI calculator, the patient's date of birth, gender, height, and weight). In instances where the user is only interested in using the BMI calculator, BP calculator, or lipid assessment, the user is asked to enter this information. For the lipid assessment, the user is asked to input the type of sample that was drawn (with response options of fasting, nonfasting, and unknown) in addition to BMI. For each of the modules, the user is automatically moved to the recommendations screen showing a patient summary of the data entered and the recommendations, after indicating that the user has finished entering the data.

The tool is a native Android and iOS app that clinicians can use on smartphones or tablet devices to evaluate patient risk, recommend healthy behaviors to prevent the development of risk, and carry out complex calculations to determine the appropriate interventions, as recommended by the NHLBI guidelines. The development of the app content [7,8], implementation protocol [10], and results of an 18-month, cluster randomized trial in 32 clinical practices are described elsewhere [13].

Figure 1. Screenshots illustrating (from left to right) (1) the 4 modules available at login: integrated risk assessment, body mass index (BMI) and blood pressure (BP) calculators, and lipid management instrument; (2) the BP calculator input screen; (3) the BMI calculator input screen; and (4) the lipid management input screen. HDL-C: high-density lipoprotein cholesterol; TC: total cholesterol; TG: triglycerides.



Approach

We used an iterative process of designing, testing, and revising throughout the design and development life cycle for the screening instrument, the validated BMI and BP calculators, and the lipid assessment instrument [7]. Pretesting, conducted after initial development of the app, was completed in 2 phases. First, we conducted one-on-one in-person testing with clinicians. Second, at a subsequent phase of testing, we examined the use of the app in clinical practice. Using a quantifiable instrument, we also asked the clinical practice testers about their overall experience with the CDS tool. Both testing cycles were reviewed by RTI International's Institutional Review Board and deemed exempt because no personally identifiable information was obtained from participants and the data gathered were used for systems research.

In-Person Testing

Based on current recommendations from evidence-based user-experience research [16,17], we initiated in-person usability testing by recruiting a convenience sample of 5 clinicians from 2 local universities in the Raleigh-Durham, North Carolina area, using a snowball recruiting approach. A member of the research team conducted one-on-one testing, which lasted for approximately 1 hour. Each participant was given a brief overview of the app and 4 test cases to complete—one each for the screening instrument and the 3 validated calculator modules that the app supports. Test cases were authored by 1 of the authors and CDS product owner (RDF) with domain expert oversight by another of the authors (KAL), which were then reviewed and approved by the chair of the NHLBI Expert Panel. Together, the test cases (Figure 2) represent the clinical scenarios and essential frequent tasks for which the tool would be used. Participants were instructed to enter the data as shown in the scripted scenario using a test device without assistance and to "think aloud" as they went, as is typical in usability testing [18]. The testing sessions were audio-recorded, capturing participants' verbal feedback. We did not offer participants financial incentives for their participation.

Testing in Clinical Practice

Because the CDS tool is designed primarily for use by pediatricians, we recruited participants for in-clinic usability testing through the American Academy of Pediatrics (AAP) Quality Improvement Innovation Network member listserv. Members of this professional group are particularly relevant to sample because participants are board-certified pediatricians who are specifically interested in testing practical tools that improve the quality of care for children and their families. A total of 29 clinicians responded to the recruitment email—which invited clinicians to attend a half-hour webinar detailing the project and CDS app—and agreed to use the app in at least 5 patient-encounter scenarios during a 2-week period. We did not offer any financial incentives to clinicians for their participation. We made the tool available to all interested participants through TestFlight, a Web service through which the research team managed access to prerelease versions of the app. Among the 29 initial clinician responders, 19 clinicians downloaded the app and 14 clinicians provided feedback on their user experiences over the 2-week testing period. We gave participants the opportunity to provide immediate, asynchronous feedback on their user experience in the form of unstructured comments via email, telephone, or short message service during the testing period rather than waiting until the end. We included the telephone number and email address of a research team member in the "About" section of the app so that participants could easily provide such feedback. On completion of testing, we instructed participants to delete the app.

A total of 14 participants completed the 10-item System Usability Scale (SUS) [19] questionnaire, which was administered electronically (Textbox 1), to capture their overall experience with the app. To calculate the SUS composite score of the overall usability, we summed the score contributions from each item, which ranged from 1 to 5. For items 1, 3, 5, 7, and 9, the score contribution is the scale position (eg, 4=agree) minus 1. For items 2, 4, 6, 8, and 10, the contribution is 5 minus the scale position (to account for the negative phrasing of these questions). We then multiplied the sum of the scores by 2.5 to obtain the overall SUS composite score, with a possible range from 0 to 100 [19].

Figure 2. Example of a test case. BMI: body mass index; CVD: cardiovascular disease.

Risk Assessment: Case 1

Data entry element/Question	Test case response
Date of Birth	1/23/1996
Gender	Male
Height	6 ft
Weight	203 lbs
Categorize change in BMI	Stable
Blood Pressure - Systolic	134
Blood Pressure - Diastolic	60
Does the patient have any of the following conditions?	Hypertension
Categorize change in condition	Not improving
Does the patient have any other high or moderate risk conditions?	No
Categorize physical activity level	Unknown
Does this represent a decrease in physical activity level?	Unknown
Does the patient currently smoke?	No
Does the patient have a smoke free home?	No
Is either parent obese?	Yes
Does either parent have dyslipidemia?	No
Parent Test Case 1	167
Parent Test Case 2	Unknown
Does the patient have a positive family history of CVD?	Yes

Textbox 1. Items in the System Usability Scale (response options ranged from 1=strongly disagree to 5=strongly agree).

1. I think that I would like to use this clinical decision support (CDS) app frequently.
2. I found the CDS app unnecessarily complex.
3. I thought the CDS app was easy to use.
4. I think that I would need the support of a technical person to be able to use this CDS app.
5. I found the various functions in this CDS app were well integrated.
6. I thought there was too much inconsistency in this CDS app.
7. I would imagine that most people would learn to use this CDS app very quickly.
8. I found the CDS app very cumbersome to use.
9. I felt very confident using the CDS app.
10. I needed to learn a lot of things before I could get going with this app.

In addition to the composite score, we assessed the learnability and usability subscales by analyzing the average responses to learnability (questions 4 and 10 in Textbox 1) and usability items (the remaining 8 questions in Textbox 1) [20]. The learnability and usability subscale scores each have a possible range from 1 to 5.

We also asked users to respond to a few additional questions, including (1) an open-ended question asking "How many times did you use the application during a patient encounter?"; (2) "In what type of patient visit did you use the application?"—with response options of well-child visit, sports physical, weight or obesity follow-up, BP follow-up, lipid follow-up, and other (please specify); and (3) "Which component of the application did you use?"—with response options of integrated assessment, lipid assessment, BP calculator, BMI calculator, and none.

Data Analysis

We examined the feedback from in-person and clinical practice testers to identify patterns and relevant themes, with the aim of gathering actionable suggestions to revise the CDS tool. We entered the qualitative data into a matrix that segmented clinicians' comments by comment type, such as positive

comment or suggestion for improvement. This common technique in qualitative research enables researchers to organize data to identify commonalities and variations that emerge in comments [21,22]. In this study, 2 of the authors (PAW and RDF) independently reviewed comments to identify relevant patterns and themes, using a coding scheme developed by 1 of the authors (PAW). We coded comments from the in-person and clinical practice testers as being a positive comment, a negative comment, or a suggestion for improvement or change. The same author (PAW) reviewed any discrepancies in interpretation and together 2 of the authors (PAW and RDF) made a final determination.

Results

In-Person Testing

A total of 5 clinicians (4 outpatient and 1 inpatient) participated in testing the CDS tool. Each clinician had been actively practicing for more than 2 years. All participants practiced at large academic medical centers. The testers were either general pediatricians, internal medicine physicians who saw a large number of pediatric patients, or pediatricians with subspecialties in nephrology or endocrinology.

Most of the testers' comments conveyed during the think-aloud sessions involved suggestions for improvements or changes; a few comments were simply compliments, with 9% (5/58) positive comments; and there were no (0%) negative comments. Thematic analysis of the comments from the in-person testers showed that suggestions for improvements or changes fell into 3 categories: (1) layout, navigation, and/or the user experience (41%, 24/58); (2) content refinement (41%, 24/58); or (3) technical deployment (9%, 5/58).

Suggestions for changes to the app related to the user experience and reflected individual preferences for the default display and functionality, such as:

> *Does not like that the alphabet is the default keyboard, would prefer number pad.*
>
> *Default to open for the recommendations and supportive action.*

Content-related suggestions for changes to the app pointed out areas that needed clarification, particularly related to the information conveyed in the recommendations regarding Estimated Energy Requirement (EER) presented in the patient summary and NHLBI recommendations in the obesity risk section.

> *Not sure what the EER is. Would like to calculate that in the app.*
>
> *EER, a pediatrician may not know what this means.*

For children with out-of-range BMIs, we provided EER language as part of the recommendations. However, whereas nutritionists understand EER, clinicians typically do not. Consequently, we provided support terminology as a design enhancement. Nonactionable user input included specific criticism of the guideline content, which we were not at liberty to revise; for example:

> *Shorten the Overweight/Obesity Recommendations.*
>
> *Consolidate the Tobacco Exposure recommendation with Family History Recommendations.*

The few deployment suggestions provided meaningful ideas to improve the usability of the CDS tool for clinicians:

> *Would like to be able to print physical activity and nutrition/diet recommendations for the patient to take home.*
>
> *Would like to email the patient [the] patient summary, activity, diet and personal risk factor information.*

However, some user input was not actionable because it conflicted with the intended design for a freestanding application; for example:

> *This information is redundant to the information available in/entered in the EHR [electronic health record].*

When considering the screening instrument and the 3 validated calculator modules as well as the functions of the tool, the majority of suggestions for improvement (59%, 34/58) were in response to the content and display recommendations provided by the app based on the patient's risk factor information input by the user. Common themes included the length of text displayed and the formatting of text, such as:

> *Supportive action in Family History should only be displayed one time.*
>
> *Change blood pressure to number spinner rather than number data entry.*

Additionally, 28% (16/58) of comments were suggestions for improvement that could be applied to all modules in the app and spoke to the importance of keeping the interface consistent:

> *When in landscape, the font size changes.*
>
> *Would like to see metric and standard units displayed on the same screen without having to toggle between the two.*

A small number of comments related to only 1 specific module: only 9% (5/58) of the comments related to the integrated assessment and only 1 comment each (<1%) related to the BMI, BP, and lipid modules.

Testing in Clinical Practice

All 14 in-clinic testers were actively practicing pediatricians whose patient population was more than 80% pediatric. Although 2 participants elected to provide limited, asynchronous feedback via email and telephone during the testing period, most of the input was submitted by all clinicians after completion of the 2-week testing period. The app was used between 1 and 20 times per participant during the testing period, with an average of 7 uses. Participants reported using the app most frequently in well-child visits (87%, 13/15), followed by weight or obesity follow-ups (53%, 8/15), BP follow-ups (40%, 6/15), sports physicals (40%, 6/15), lipid follow-ups (20%, 3/15), and other types of visits (20%, 3/15). The most commonly used modules in the app were the integrated assessment (86%, 13/15) and BP calculator (86%, 13/15), followed by the BMI calculator (73%, 11/15), lipid assessment (47%, 7/15), and none

(0%). The results of the SUS data analysis showed that general reactions toward the CDS tool were positive, given the average score of 81, meaning the app was viewed as "above average" with respect to usability (as defined by a score >68; [20]). The learnability subscale average of 1.53 showed that most participants did not think they would need to learn more or require technical support to use the app. Additionally, the 8-item usability subscale average of 3.36 showed that most participants rated the usability of the CDS tool favorably [20].

Consistent with the results from the SUS, general reactions toward the CDS tool were positive, with 35% (6/17) positive comments and no (0%) negative comments; for instance:

> *Great app - would use it at most visits.*

Users particularly appreciated the app's feature of calculating the BP percentile based on the NHLBI's BP tables for children and adolescents [23]:

> *Overall very easy to use and helpful to not have to look up BP values on the chart.*

Thematic analysis of the comments from the clinical practice testers showed that suggestions for improvements or changes fell into 2 categories: (1) layout, navigation, and/or the user experience (24%, 4/17), or (2) content (41%, 7/17). Suggestions for changes to the app related to the user experience and to changing the app to give the user control and empower the user, especially clinicians using the app on the iPhone:

> *When entering numbers on the iPhone, I had to click past the alpha keyboard to get to the numbers. Other apps I have used have the number keypad come up first!*
>
> *The touch screen did not respond easily to touch and many features were very erratic in their scrolling, such as dates.*
>
> *The associated algorithms are great but difficult to properly visualize on the small screens of the smartphones.*

Content-related suggestions for changes to the app focused on the amount of information conveyed in the recommendations. Sometimes users thought the app provided too much information and they suggested reorganizing the information to make it more succinct:

> *The recommendations should be narrowed down using the answers entered. Otherwise there are too many and it becomes cumbersome to use.*
>
> *The recommendations were too lengthy to be useful in a clinical visit.*
>
> *The recommendation sat the end of the assessment are very wordy. It is a lot of information and the recommendations are important, so streamlining that will mean more people use the app.*

Nonactionable user input included specific criticism of the guideline content, which we were not at liberty to revise, for example:

> *My only problem with the app was it gave too many repetitive recommendations for healthy children with no risk factors. For example, I would enter data for a child with 25% BMI, enter nonsmoking for child and parents, normal values for parent cholesterol values and normal answers to questions about family risk factors, and it would still recommend I ask about smoking and family risk factors after that.*

In other instances, testers wanted additional information and information better tailored to the individual patient:

> *I really wanted to know how to intervene when I had abnormal lipids, but that wasn't easily accessible.*
>
> *During the integrated assessment it seemed to give the general guidelines more than telling what to do with this specific patient.*

Discussion

Principal Findings

Clinical decision support tools can improve clinician adherence to guidelines [11,13]. This paper demonstrates the process and value of testing and adapting a CDS tool to assist clinicians in implementing the NHLBI "Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents" [3], using a user-centered design approach. Testers generally responded positively toward the CDS tool. They particularly appreciated the app's feature of calculating the BP percentile based on the NHLBI's BP tables for children and adolescents. Testers suggested changes to the app related to user experience, content refinement, and technical deployment. The majority of the suggested changes centered on the content and display of the recommendations for clinicians, including making the app more user-friendly for clinicians using the app on the iPhone and reorganizing and tailoring the recommendations. These findings are similar to those from other usability studies of decision support tools, which often show that testers recommend clearer, more concise content; a more user-friendly layout design; and improvements in navigation [15,24,25] to enhance tools.

However, we employed a user-centered approach and systematic process to develop the Pediatric Cardiovascular Risk Reduction CDS Tool that many developers do not implement. Our approach illustrates a field-proven method for soliciting expert user input from a geographically distributed sample of difficult-to-reach participants. Collaboration with the AAP, the credentialing body for all board-certified pediatricians in the United States, enabled access to members of the AAP's Quality Improvement Innovation Network. This network was established to provide a standard mechanism for developing practical and usable measures, tools, and strategies for the practicing pediatrician in a primary care practice as well as the pediatric hospitalist in the inpatient setting. Engagement with this professional association provided a point of entry to a practical working laboratory for gathering pediatrician input based on their use of the CDS in real-world patient encounters.

We incorporated tester feedback into the Pediatric Cardiovascular Risk Reduction CDS Tool design when feasible and applicable. Actionable user input focused mainly on matters of user experience and recommendations to streamline the use

of the app during clinical visits. We received feedback that translated into changes in data entry, presentation of recommendations, and presentation of critical results. For example, testers recommended simplifying data entry, which resulted in our asking fewer questions and implementing persistence of forms-based variables across all of the instruments, meaning data entered in 1 tool would carry over to another.

Testers also recommended reducing the volume of results displayed, which informed the restructuring of the recommendations layout using design patterns. Consequently, we used faceted navigation and presentation of information to provide an integrated, incremental search-and-browse experience to increase tailoring and refinement of the results presented for an individual patient. The more information—such as social and family history, known risk factors, and clinical observations—that the clinician enters in the CDS tool, the more facets are used progressively to refine results, eliminating extraneous information by narrowing search results. This empowers the user by not forcing clinicians to enter data for all of the variables on a screener page. Consistent with user input regarding the burden of data entry, we did not force data entry and enabled clinicians to figure out how much data they wanted to enter in the screener.

User input on the presentation of recommendations also led to refinements to improve readability of the content on the summary results page. Consequently, we reduced the amount of text and used input from testers to make the recommendation language more actionable (eg, “Measure fasting lipid panel 2x and average results”) rather than providing a more detailed recommendation for follow-up.

Additionally, suggestions led to us providing immediate user feedback with the addition of a color-coding feature (red, yellow, green) to highlight critical or elevated results, such as highlighting when out-of-range or borderline results emerged for BMI, BP percentile, and lipids. Finally, nonactionable user input included specific criticisms of the guideline content, health information technology deployment, and workflow issues, all of which were beyond the scope of this CDS design.

Limitations

One limitation of the study is that we tested the tool using clinicians who were interested in quality improvement and motivated to adopt decision aids. Consequently, participating clinicians may have been more readily amenable to using the CDS tool or more adept at using it than clinicians without this background, experience, or interest. This, and other individual factors that we did not examine (eg, age), may affect the way clinicians adopt the CDS tool. Certainly testing with more “typical” clinicians who may be less technology savvy would be valuable in future work to identify and address as many usability issues as possible and to ensure that the app is user-friendly for those less familiar with the technology.

Another limitation is that the CDS tool was a part of a multifaceted intervention, which limited our ability to assess individual clinician engagement with the app in more detail. During our testing in clinical practice, we did not measure the duration of time spent with the app during the patient visit. Further, we did not ask clinicians to comment on how the CDS tool affected the patient visit and the patient-provider relationship during the clinical practice testing, which would ultimately be a contributing factor to whether or not the CDS is adopted in the clinical setting.

We also were unable to retest the refined design after incorporating user feedback because of resource constraints. However, the user feedback described in this paper was applied to changes made in the final version of the Pediatric Cardiovascular Risk Reduction CDS Tool and testers were informed of how their input from usability testing had been incorporated in the final build.

Additionally, for the development of this tool, we were responding to a request for a freestanding decision support application. Consequently, we deliberately built it in the context of physician maintenance of certification criteria and to support the trend of individual physicians and staff to use personal mobile phones or other devices. The advantage of this is that the tool is highly portable, which makes it easier to use within the clinical workflow. However, the lack of integration with electronic health records (EHRs) requires additional effort to enter relevant clinical information. This may limit the uptake of this freestanding tool. As EHRs become more prevalent, integration will likely become more of an issue. In the future, developers should consider the relative benefit of building a Web application with an application program interface library, including clear standards for the exchange of clinical variables and bidirectional communication functions between the decision support application and other clinical information systems. Future work should also concentrate on how the CDS will integrate into the broader health care ecosystem.

Finally, while this paper focused on the usability testing, rather than the implementation and effectiveness (see [9,10,13] for details on these aspects) of the CDS tool, it is worth noting that questions remain about the specific components of CDS tools that are effective, the impact of CDS tools on patient outcomes and clinical workload, and clinician preferences for certain CDS features [26-29]. This CDS tool was intended to improve guideline uptake. Whereas prior research has shown that many CDS tools improve clinician adherence to guidelines [11] and other aspects of their performance, the effects of such tools on patient, economic, workload, and efficiency outcomes are understudied [26,27]. However, the Pediatric Cardiovascular Risk Reduction CDS Tool—in combination with the other tools, education, and support that comprised the full comprehensive, multifaceted intervention—was effective in improving adoption of the guidelines [13]. Improved patient outcomes should follow from clinician implementation of these guidelines, if the patient implements the suggested behavior changes. Clinical decision support tools that reduce clinicians’ efforts to digest and impart recommendations have been shown to be central to improving patient care [28], which was a primary benefit of this tool that focused on making a massive set of NHLBI guidelines accessible to clinicians. Other features of successful CDS tools that our CDS tool did not incorporate include providing advice for patients and clinicians at the same time to support improved patient-provider communication or shared decision making and

requiring clinicians to give reasons when overriding CDS recommendations [29].

Conclusions

This study assessed the usability of the Pediatric Cardiovascular Risk Reduction CDS Tool by clinicians. Testing through both one-on-one in-person testing using a "think aloud" approach and in-practice use of the tool along with formal usability metrics revealed ways to optimize the tool related to the user experience, content, and deployment. Although this paper focuses on a CVD tool, the insights that we shared about the reactions to testing and adapting this tool may be applicable to the design of other mobile health apps and CDS tools. Future work should bear in mind the benefits of integration with EHRs as they become more prevalent in the coming years.

Acknowledgments

This study was funded by the National Heart, Lung, and Blood Institute, NHLBI-PB-(HL)-2009 195D HHSN268200900120U, and by the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Labarthe DR, Dunbar SB. Global cardiovascular health promotion and disease prevention: 2011 and beyond. Circulation 2012 May 29;125(21):2667-2676 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.087726] [Medline: 22644371]
2. Hayman LL, Helden L, Chyun DA, Braun LT. A life course approach to cardiovascular disease prevention. Eur J Cardiovasc Nurs 2011 Jul;10 Suppl 2:S20-S31. [doi: 10.1016/S1474-5151(11)00113-7] [Medline: 21762848]
3. National Heart Lung Blood Institute (NHLBI). Washington, DC: Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute; 2012. Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents: Full report. (12-7486) URL: http://www.nhlbi.nih.gov/files/docs/guidelines/peds_guidelines_full.pdf [accessed 2016-05-26] [WebCite Cache ID 6hnOl5JTo]
4. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999 Oct 20;282(15):1458-1465. [Medline: 10535437]
5. Hagan JF, Hagan JS, Duncan PM. Bright Futures: Guidelines for Health Supervision of Infants, Children, And Adolescents. 3rd edition. Elk Grove Village, IL: The American Academy of Pediatrics; 2008.
6. Martin LA, Ariza AJ, Thomson JS, Binns HJ, Pediatric Practice Research Group. Seconds for care: evaluation of five health supervision visit topics using a new method. J Pediatr 2008 Nov;153(5):706-11, 711.e1. [doi: 10.1016/j.jpeds.2008.05.001] [Medline: 18589443]
7. Furberg R, Bagwell J, LaBresh K. Development of the pediatric cardiovascular risk reduction clinical decision support tool. 2013 Presented at: mHealth Summit; 2013; Washington, DC.
8. Kish-Doto J, Moultrie R, McCormack L, Furberg R, LaBresh K. Assessing patient-provider communication barriers to implementing new expert panel risk reduction guidelines. J Commun Healthc 2014;7(3):214-227.
9. LaBresh KA, Lazorick S, Furberg RD, Ariza A, Whetstone L, Hobbs C, et al. Improvement of guideline use in young hearts strong starts, a cluster randomized trial of cardiovascular prevention in pediatric offices. Circulation 2013;128:A13865.
10. LaBresh KA, Lazorick S, Ariza AJ, Furberg RD, Whetstone L, Hobbs C, et al. Implementation of the NHLBI integrated guidelines for cardiovascular health and risk reduction in children and adolescents: rationale and study design for young hearts, strong starts, a cluster-randomized trial targeting body mass index, blood pressure, and tobacco. Contemp Clin Trials 2014 Jan;37(1):98-105 [FREE Full text] [doi: 10.1016/j.cct.2013.11.011] [Medline: 24295879]
11. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S. Computerized clinical decision support for prescribing: provision does not guarantee uptake. J Am Med Inform Assoc 2010 Jan;17(1):25-33 [FREE Full text] [doi: 10.1197/jamia.M3170] [Medline: 20064798]
12. Schuler D, Namioka A. Participatory Design: Principles and Practices. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1993.
13. LaBresh KA, Ariza AJ, Lazorick S, Furberg RD, Whetstone L, Hobbs C, et al. Adoption of cardiovascular risk reduction guidelines: a cluster-randomized trial. Pediatrics 2014 Sep;134(3):e732-e738 [FREE Full text] [doi: 10.1542/peds.2014-0876] [Medline: 25157013]
14. Furberg R, Bagwell J, Kizakevich P, Eckhoff R, Zhang Y, Bakalov V, et al. A mobile clinical decision support tool for implementing the NHLBI Expert Panel Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescent. 2013 Presented at: mHealth Summit; 2013; Washington, DC.
15. Payne AY, Surikova J, Liu S, Ross H, Mechetiuc T, Nolan RP. Usability Testing of an Internet-Based e-Counseling Platform for Adults With Chronic Heart Failure. JMIR Human Factors 2015 May 08;2(1):e7. [doi: 10.2196/humanfactors.4125]
16. Nielsen J. 2000 Mar 19. Why You Only Need to Test with 5 Users URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users [accessed 2016-04-25] [WebCite Cache ID 6h29pu9Jn]

17. Nielsen J. 2012 Jun 04. How Many Test Users in a Usability Study? URL: https://www.nngroup.com/articles/how-many-test-users/ [accessed 2016-04-25] [WebCite Cache ID 6h2ANP9cL]
18. 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 Human Factors 2015 Sep 10;2(2):e14. [doi: 10.2196/humanfactors.4537]
19. Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan P, editor. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996:189-194.
20. Lewis J, Sauro J. The Factor Structure of the System Usability Scale. In: Kurosu M, editor. Human centered design. Berlin: Springer; 2009:94-110.
21. Gribbs G. Analyzing Qualitative Data. London, UK: Sage Publications Ltd; 2007.
22. Miles M, Huberman M. Qualitative Data Analysis. 2nd edition. Thousand Oaks, CA: SAGE Publications; 1994.
23. National Heart, Lung, and Blood Institute (NHLBI). The Fourth Report on the Diagnosis, Evaluation, and Treatment of High Blood Pressure in Children and Adolescents. 2005 May. Blood Pressure Tables for Children and Adolescents URL: https://www.nhlbi.nih.gov/files/docs/resources/heart/hbp_ped.pdf [accessed 2016-05-27] [WebCite Cache ID 6hotOVIUX]
24. Berry DL, Halpenny B, Bosco JL, Bruyere J, Sanda MG. Usability evaluation and adaptation of the e-health Personal Patient Profile-Prostate decision aid for Spanish-speaking Latino men. BMC Med Inform Decis Mak 2015 Jul;15:56 [FREE Full text] [doi: 10.1186/s12911-015-0180-4] [Medline: 26204920]
25. Li LC, Adam PM, Townsend AF, Lacaille D, Yousefi C, Stacey D, et al. Usability testing of ANSWER: a web-based methotrexate decision aid for patients with rheumatoid arthritis. BMC Med Inform Decis Mak 2013;13:131 [FREE Full text] [doi: 10.1186/1472-6947-13-131] [Medline: 24289731]
26. Bright TJ, 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]
27. Garg AX, Adhikari Neill K J, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
28. Kawamoto K, Houlihan C, Balas E, Lobach D. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Br Med J 2005;330(7494):765-768.
29. Roshanov PS, 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. BMJ 2013 Feb 14;346:f657. [doi: 10.1136/bmj.f657]

Abbreviations

AAP: American Academy of Pediatrics
BMI: body mass index
BP: blood pressure
CDS: clinical decision support
CVD: cardiovascular disease
EER: Estimated Energy Requirement
EHR: electronic health record
NHLBI: National Heart, Lung, and Blood Institute
SUS: System Usability Scale

Edited by G Eysenbach; submitted 21.12.15; peer-reviewed by A Payne, A Press, A Roundtree; comments to author 24.01.16; revised version received 03.03.16; accepted 10.05.16; published 21.06.16.

Please cite as:
Williams PA, Furberg RD, Bagwell JE, LaBresh KA
Usability Testing and Adaptation of the Pediatric Cardiovascular Risk Reduction Clinical Decision Support Tool
JMIR Hum Factors 2016;3(1):e17
URL: http://humanfactors.jmir.org/2016/1/e17/
doi:10.2196/humanfactors.5440
PMID:27328761

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org