JMIR Human Factors

Impact Factor (2023): 2.6
Volume 5 (2018), Issue 2 ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

Contents

Original Papers

Viewpoint

Review

Original Paper

Designing for Clinical Change: Creating an Intervention to Implement New Statin Guidelines in a Primary Care Clinic

Melissa DeJonckheere[1], PhD; Claire H Robinson[2], MPH; Lindsey Evans[2], BSN, MPH, MPP, RN; Julie Lowery[2], MHSA, PhD; Bradley Youles[2], MPA; Adam Tremblay[3,4], MD; Caitlin Kelley[2], MSI; Jeremy B Sussman[2,3,5], MS, MD

[1]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

[2]Center for Clinical Management Research, Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, MI, United States

[3]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[4]General Medicine, Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, MI, United States

[5]Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Jeremy B Sussman, MS, MD
Center for Clinical Management Research
Veterans Affairs Ann Arbor Healthcare System
2215 Fuller Road
Ann Arbor, MI, 48105
United States
Phone: 1 734 845 3502
Email: jeremysu@med.umich.edu

Abstract

Background: Recent clinical practice guidelines from major national organizations, including a joint United States Department of Veterans Affairs (VA) and Department of Defense (DoD) committee, have substantially changed recommendations for the use of the cholesterol-lowering statin medications after years of relative stability. Because statin medications are among the most commonly prescribed treatments in the United States, any change in their use may have significant implications for patients and providers alike. Prior research has shown that effective implementation interventions should be both user centered and specifically chosen to address identified barriers.

Objective: The objectives of this study were to identify potential determinants of provider uptake of the new statin guidelines and to use that information to tailor a coordinated and streamlined local quality improvement intervention focused on prescribing appropriate statins.

Methods: We employed user-centered design principles to guide the development and testing of a multicomponent guideline implementation intervention to improve statin prescribing. This paper describes the intervention development process whereby semistructured qualitative interviews with providers were conducted to (1) illuminate the knowledge, attitudes, and behaviors of providers and (2) elicit feedback on intervention prototypes developed to align with and support the use of the VA/DoD guidelines. Our aim was to use this information to design a local quality improvement intervention focused on statin prescribing that was tailored to the needs of primary care providers at our facility. Cabana's Clinical Practice Guidelines Framework for Improvement and Nielsen's Usability Heuristics were used to guide the analysis of data obtained in the intervention development process.

Results: Semistructured qualitative interviews were conducted with 15 primary care Patient Aligned Care Team professionals (13 physicians and 2 clinical pharmacists) at a single VA medical center. Findings highlight that providers were generally comfortable with the paradigm shift to risk-based guidelines but less clear on the need for the VA/DoD guidelines in specific. Providers preferred a clinical decision support tool that helped them calculate patient risk and guide their care without limiting autonomy. They were less comfortable with risk communication and performance measurement systems that do not account for shared decision making. When possible, we incorporated their recommendations into the intervention.

Conclusions: By combining qualitative methods and user-centered design principles, we could inform the design of a multicomponent guideline implementation intervention to better address the needs and preferences of providers, including clear and direct language, logical decision prompts with an option to dismiss a clinical decision support tool, and logical ordering of feedback information. Additionally, this process allowed us to identify future design considerations for quality improvement interventions.

(JMIR Hum Factors 2018;5(2):e19) doi:10.2196/humanfactors.9030

KEYWORDS

cardiovascular disease; preventative medicine; clinical decision support; user-centered design; qualitative research; implementation

Introduction

Background

There has been a dramatic shift with respect to how guidelines recommend that American medical providers should prescribe commonly used cholesterol-lowering statin drugs [1]. In 2013 and 2014, the American College of Cardiology (ACC) and American Heart Association (AHA) and the United States Department of Veterans Affairs (VA) and Department of Defense (DoD) released new clinical practice guidelines on the treatment of blood cholesterol to reduce cardiovascular risk in adults [2-4]. Where previous guidelines had focused on giving increasing doses of statins until a patient's cholesterol level dropped below a specific target, both new guidelines recommend fixed doses of medicine based on the patient's atherosclerotic cardiovascular disease (ASCVD) *risk*, the chance that the patient will develop cardiovascular disease (CVD) [2,5-9].

The new guidelines present challenges to adoption. First, moving away from cholesterol target-based treatment models represents a conceptual change in clinical practice. Second, in many cases, risk-based guidelines would require providers to calculate a patient's ASCVD risk, which could substantially alter a provider's workflow. Third, the discrepancies between the VA/DoD and ACC/AHA guidelines could cause confusion [10].

Effective implementation of new guidelines should recognize the existing barriers to adoption [11], including providers' knowledge, attitudes, and behaviors about those guidelines [12]. They must also address those barriers in ways that are effective, accurate, and user centered. Existing strategies, including provider education, clinical decision support, and audit and feedback must address the barriers and the providers' needs [12]. This requires a strong framework for designing an intervention and for making sure the intervention is effective.

Effective interventions fit the needs of the end users. To this end, user-centered design focuses on understanding the physiological, cognitive, and social aspects of the intended user that could alter how someone will use a tool or system [13]. In a health care setting, user-centered design can be employed to create or adapt tools that are consistent with the physiological, cognitive, and social needs of providers to address challenges to adoption and increase the likelihood of their use.

Objectives

In this study, we developed and tested a multicomponent guideline implementation intervention (hereafter referred to as the intervention) to improve statin prescribing. Our intervention was developed with semistructured qualitative interviews, an established theoretical framework, and principles of user-centered design. This paper describes the intervention development process with providers, which was conducted to (1) illuminate knowledge, attitudes, and behaviors and (2) elicit feedback on intervention prototypes developed to align with and support the use of the VA/DoD guidelines. Our aim was to use this information to design a local quality improvement intervention focused on statin prescribing that was tailored to the needs of primary care providers.

Methods

Intervention Background and Development

We sought to develop and test a multicomponent guideline implementation intervention to improve statin prescribing. The processes were designed to find new, but practical, components for the intervention and help us improve the processes that we already had planned based on the literature and existing practices. For example, research demonstrates that a multicomponent intervention is often more effective than a single approach alone [14].

In the following section, we describe the components of the intervention: educational program, clinical decision support tool, and performance measurement with audit and feedback. Paper-based prototypes were created as working models to be tested for acceptability before investing in computerized systems. The prototypes were modeled after currently existing tools utilized in the VA health system to facilitate providers' ability to imagine how the prototypes would function in their current workflow. In our user-centered design process, we asked providers to identify their needs and preferences specific to the 4 prototypes described below.

Educational Program

In collaboration with providers, we developed an educational program about the new clinical practice guidelines to be delivered to all providers before the intervention began. The educational program lasted 15 min. It included a summary of the guidelines and 3 cases that demonstrated the differences between the new guidelines and the old. We also developed a single-page tool describing and comparing the VA/DoD and the ACC/AHA Clinical Practice Guidelines (see Multimedia Appendix 1). The single-page tool was designed to be a concise and convenient reminder of changes to the statin guidelines.

Clinical Decision Support Tool

The clinical decision support tool was designed to address 2 predicted quality gaps—the traditional reminder role of pop-ups and a need to make it easier to follow the guidelines. The new guidelines require providers to calculate the risk of ASCVD for some patients using an algorithm that incorporates risk factors (eg, age, sex, smoking, high-density lipoprotein cholesterol, total cholesterol, systolic blood pressure), which significantly complicates use of the guidelines. In our facility, an ASCVD risk algorithm has not yet been incorporated into the electronic medical record, and providers access ASCVD risk calculators through external websites. The paper-based prototype resembled the existing clinical decision support alerts, thus meeting the reminder role and automatic calculations of ASCVD risk of

computerized clinical decision support (see Multimedia Appendix 2). The impact of not having the tool in the electronic health record was evaluated in our interviews.

Performance Measurement With Audit and Feedback

Traditionally, performance measurement is used in pay-for-performance programs within the VA. We worked with the VA's Center for Analytics and Reporting to create a novel performance measure that is aligned with the VA/DoD guidelines (hereafter referred to as the VA proposed performance measure; see Multimedia Appendix 3). In the VA proposed performance measure, providers would have stronger incentives for patients for whom treatment is more likely to be clinically important using a weighted point measurement system to create risk categorization of patient groups. For example, the VA proposed performance measure would award providers different points for prescribing a moderate dose statin to a patient with clinical ASCVD (5 points), a patient with diabetes (3 points), and a patient with a 10-year ASCVD risk greater than 12% (1 point). In distinction, other performance measures, such as those of the Healthcare Effectiveness Data Information Set, do not incorporate risk prediction in patient treatment recommendations. The weighted VA proposed performance measure was designed to emphasize prevention through risk calculation.

We designed an audit and feedback report template (see Multimedia Appendix 4), wherein providers would be informed of their individual performance on the VA proposed performance measure. The template includes 2 provider performance reports. The first includes breakdown of provider performance by patient risk categorization (eg, history of ASCVD; diabetes; low-density lipoprotein, LDL >190; high risk; low risk). The second displays performance by overall statin use across patients. Similar to the VA proposed performance measure, the audit and feedback report features risk prioritization.

Setting and Participants

This local quality improvement project was conducted in primary care at a single Veterans Affairs Medical Center (VAMC) between late October 2015 and June 2016. In total, 37 professionals across 5 Patient Aligned Care Teams (PACTs)—including 32 physicians with their own patient panel at the start of the project and 5 pharmacists—were invited to participate in qualitative interviews via email. Though 37 professionals were invited, data collection was designed to continue until thematic saturation was reached [15-17]. Invitees were presented with a project information sheet at the time of initial email contact, which was reviewed at the time of the interview.

Ethical Considerations

Veterans Health Administration (VHA) Handbook 1058_05 [18] provides guidance about authorization of manuscripts that have been developed through nonresearch activities (ie, without institutional review board approval under the authority of VHA operations). All VHA authors of this manuscript attest that the activities that resulted in producing this manuscript were not conducted as part of a research project but as part of the nonresearch evaluation conducted under the authority of the VA's Quality Enhancement Research Initiative.

Data Collection

We conducted semistructured interviews with providers to guide the development and testing of a multicomponent guideline implementation intervention. A qualitative approach was selected to explore user knowledge, attitudes, and behavior to improve the adaptation and implementation of the intervention. Interviews were audiotaped, transcribed verbatim, and lasted an average of 49 min. One member of the research team (CR) conducted all interviews while a research assistant took notes. Providers were not compensated for their time, and participation was completely voluntary. We began by eliciting feedback on determinants of providers' guideline uptake. We then engaged providers in a user-centered design process to examine and improve prototypes for the 4 components of the intervention.

Analysis

After a review of the literature, we determined that existing frameworks could be used to understand implementation of clinical guidelines in our setting. We used in-depth qualitative research principles structured by the Clinical Practice Guidelines Framework for Improvement [12] to guide our understanding of the barriers and facilitators to use and Nielsen's Usability Heuristics [19] to guide the user-centeredness of our development process. The Clinical Practice Guidelines Framework for Improvement examines individual-level factors (knowledge, attitudes, and behaviors) of providers [20], whereas Nielsen focuses on elements of user-centeredness and design.

Using an initial codebook based on constructs from the Clinical Practice Guidelines Framework for Improvement and Nielsen's Usability Heuristics (see Table 1), we (JS, CR, and MD) used a deductive approach to apply descriptive codes to 3 transcripts and modified our codes based on the data. We then applied codes from the modified codebook to 3 more transcripts and discussed our codes to determine consensus. The remaining 9 transcripts were coded by 1 team member (MD). We used QSR International's NVivo version 11 data analysis software to apply codes to segments of text and to create code reports that grouped all text sharing the same code. Code reports were then summarized independently by the project team members (JS, CR, MD) and discussed to reach shared understanding of themes.

Project team members (JS, CR, CK, and BY) discussed interview notes in team meetings. Following a user-centered design approach, the team discussed provider needs related to the intervention and made changes to the wording and format of the prototypes as interviews progressed. When there was misalignment between providers' preferences and design decisions, we used an adapted consensus process [21] to decide which suggested modifications were feasible.

Table 1. Initial codebook incorporating individual-level factors and elements of design.

Code	Definition
Guideline factors	
Familiarity	Awareness/knowledge/use of the guidelines
Self-efficacy	Ability to follow the guideline
Expected efficacy	Will improve clinical outcomes for patients (prevent heart attacks and strokes)
Previous practice	Change from previous care? How much does changing care affect the provider?
Use of guidelines in general	Motivated by/trust/use of external guidelines in general
Accept/reject guidelines	Agreement/disagreement with new guidelines
Risk-benefit comparison	How do the benefits to patient/outcomes compare to the risks of implementing guidelines
Evidence-based	Perception that guidelines are consistent with evidence-based practice (credibility)
Oversimplified cookbook	Concerned that the guideline is too regimented, missing real-world nuance
Autonomy	Effect on autonomy
Standardization of practice	Makes it so all providers provide similar care
Teamlet role/responsibility	Role of nursing, pharmacy, other staff in patient adherence to statins
Clarity	Ability to understand the guidelines
Gaming	Activity that produces apparent change in the measure, but no genuine change in the underlying performance
Patient factors	
Patient resistance	Willingness of patients to take medications, engage in conversation, accept recommendations
Patient tolerance	Side effects of medication prohibit adherence
Shared decision making	Effect of guidelines on shared decision making
Provider factors	
Clinical influences	Who influences uptake? Professional role, individual respect, professional, and/or personal interactions?
Performance pay	Does reimbursement or performance pay alter uptake?
Performance measurement system	Agreement with use of performance measurement system
Audit and feedback-Pt-level feedback	Use of fallout reports with specific patients to target/follow up with
Communication with patients	Strategies or tools for effective communication with patients
Practice setting factors	
Reminder system (decision support tool)	Need for a reminder system for ease of use, understanding, calculation, etc
Catch missed patients	Tool helps recognize who would benefit
# of clinical reminders	Amount of clinical reminders seen by providers
Provider education	Educational resources, strategies, tools for providers
Not applicable to practice population	Relevance of guidelines to practice
Not practical in our setting	Would require unavailable technology, nonformulary medicines, or unavailable specialists
Insufficient staff or support	Ability of practice to use guidelines with existing staff resources
Practicality/prioritization	Time to address guideline, fit with workflow
Usability heuristics	
Transparency of calculation	Provider understands how the recommendation was determined
Autonomy/allows complexity	Allows for and explains provider choices (eg, emergency exit)
Accuracy	Are the recommendations correct (by what they intend to have)
Cognitive ease of use	Saves or creates providers the need to think, calculate, remember
Speed/ease of use	Time-consuming/saving, fits workflow

Results

Participant Description

In total, 15 individuals—13 physicians and 2 clinical pharmacists—participated in interviews and represented all 5 PACT teams. In total, 9 providers did not respond to 3 email invitations and 13 declined to participate. There were no observable differences in gender, age, or participation in the educational seminar, between those we interviewed and those who did not participate.

Summary of Findings

Overall, providers were generally comfortable with the paradigm shift to risk-based guidelines but less clear on the need for the VA/DoD guidelines in specific. They preferred tools that helped them provide the care they wanted to provide without limiting their autonomy (see Table 2 for abbreviated list of changes made in response to interviews; see Multimedia Appendix 5 for detailed list of user-centered design changes).

Guidelines

Providers Accept the Paradigm Shift in Cholesterol Treatment but Some Question the Need for Separate Department of Veterans Affairs/Department of Defense (VA/DoD) Statin Guidelines

Most providers felt the risk paradigm was more closely aligned to their clinical perspective:

> *We've moved away from focusing on LDL, this one just seems more compelling…here's the person's risk, it just seems more informative and like a compelling reason to treat.* [Participant #13]

Others highlighted the benefit of providing patients with more precise, tailored risk estimates using risk-based guidelines.

One core distinction between the guidelines is that the VA/DoD guidelines are generally less aggressive than the ACC/AHA guidelines: they recommend treatment for fewer people, permit use of less-intense statin regimens, and create a gray zone where treatment is neither recommended for nor against. A few providers stated their preference for the VA/DoD guidelines and felt the ACC/AHA guidelines encouraged overtreatment. One provider explained:

> *There may be some people that are jumping right to high potency when that's not necessary, especially in the elderly population which we have a ton of.* [Participant #9]

Another participant said:

> *[I'm] not sure of the distinction between the AHA guidelines and these [VA/DoD] guidelines.* [Participant #8]

Several providers did not recognize the need for separate VA guidelines at all. One of the participants admitted:

> *...most of my colleagues here have kind of adapted it [the VA/DoD guidelines].* [Participant #15]

A few developed approaches that incorporated aspects of both sets of guidelines, such as one who appreciated the deemphasis of routine cholesterol monitoring in the VA/DoD guidelines but preferred the risk cut points established in the ACC/AHA guidelines.

Due to their patient population, a few providers noted that the differences between the 2 guidelines would likely have a very small impact:

> *They all have diabetes, many of them smoke, and they all have hypertension. A lot of them already have cardiovascular disease, so you're not really even doing a risk assessment. Many of them don't specifically fall into the scope of this, so to be honest I haven't used the VA one much just because there's not been much need for it in the patients that I see.* [Participant #7]

Within the Risk Paradigm, Providers Are Not Confident in How to Deal With Shifting Risk

Providers were generally comfortable with the role of risk prediction in the guidelines. They did express some confusion about how to address changing risk factors and the lack of consistency of risk prediction. For these problems, they felt that the guidelines were not responsive. One provider explained:

> *I think the calculators can vary a lot, depending on what someone's blood pressure is that day or their smoking status. Those kinds of things can change. Then someone if they quit smoking might not be, you know, the same risk as they were 10 minutes ago...So, I think it's not exactly clear cut...* [Participant #14]

The Paradigm Shift Creates New Responsibilities in Doctor-Patient Communication

Several providers felt that their patients might find it difficult to shift away from cholesterol treatment targets. Patients are familiar with recommendations to improve their cholesterol numbers. As one provider explained:

> *I think there is still a little bit of resistance. Patients are really caught up on the LDL number because, I guess we used to really drive that hard, like "Oh, your LDL should be this and it's too high and so we're going to add these other drugs, or increase the dose," or whatever it might be. I think some people were still really hung up on those numbers.* [Participant #7]

Comparatively, risk reduction is more "abstract" than LDL reduction. Several providers described that patients "like to see that [the treatment is] doing something," which is difficult to demonstrate under guidelines that do not specifically call for routine cholesterol monitoring. Thus, providers were concerned about nonadherence:

> *I worry that we are going to have even more trouble initiating and getting people to adhere to statins when we are talking about them in this new kind of abstract confusing way for patients. I have probably not been*

as aggressive in moving towards these newer guidelines in part for that reason. [Participant #2]

Table 2. Abbreviated list of user-centered design considerations for intervention components.

Tool and user suggestion	Impact on adoption
Clinical decision support design	
Include high/medium/low-risk language in reminder—facilitates conversation with patient	Implemented
Disable reminder for patients receiving palliative care	Implemented
Prepopulate risk score automatically within reminder	Future consideration
Alert only when appropriate (disable reminder for patients with complicated clinical situations)	Future consideration
Add specific risk percentage in reminder rather than high/medium/low language	Not used
Add additional line for comments	Not used
Audit and feedback design	
Organize patient fallout by risk category	Implemented
Clarify provider comparison group (local vs Department of Veterans Affairs)	Future consideration
Devise mechanism/algorithm that accounts for complicated patients in performance measure and subsequently in the audit and feedback report	Future consideration
Provide credit for shared decision making	Future consideration
Include specific and actionable performance improvement suggestions	Future consideration
Remove provider percentile altogether because it creates undue angst	Not used

Clinical Decision Support Tool

Providers Desire Clinical Decision Support Tools That Allow for Cognitive Ease of Use and Speed

Providers' interest in having a clinical decision support tool during the patient encounter was based around efficiency:

> *If the reminder already calculated the risk, I'd love that. I hate having to go to the internet, or look on my smartphone, so I think the ideal reminder would calculate the risk for you.* [Participant #6]

A few providers indicated that the clinical decision support tool may be especially useful in patients whom the calculator estimates to be at high risk for ASCVD but have no history of heart attack or stroke:

> *In this particular case, I like it because this is one that may not jump out immediately at you. This person doesn't have coronary disease so it's kind of helping you work through and reminding you where the guidelines stay.* [Participant #9]

Providers Want Clinical Decision Support Tools That Allow for Autonomy

When asked about the computerization of clinical decision support tools, most providers indicated a need for autonomy within the system, whereby providers can exit or cancel a clinical reminder when it is inappropriate or inaccurate for the particular visit or patient:

> *Sometimes it seems like things come up that aren't supposed to, or they don't come up and they should…I think there's often circumstances where it's like, "How do you get out of this loop?" where this isn't right and it should go away, but you can't make it go away and so I like that there's an option for like, "This is wrong," and so you can get out of that.* [Participant #7]

Providers said they generally appreciate being reminded when a patient is not meeting a guideline but want to be able to accurately state why the patient is not on a statin rather than bend the truth simply to disable the reminder.

Providers Want Clinical Decision Support Tools That Can Be Disabled

Providers wanted a clinical decision support tool that would not continue to alert after an issue has been addressed. However, there was some debate as to which clinical situations should lead to a reminder being disabled indefinitely and which would warrant a revisited conversation:

> *If you had a discussion with the patient and they decided against it, okay, if you had a discussion with the patient and they decided for it, okay. I'd never not do it because they were poor in the past, you know, we'd have a discussion and in that moment, I'd give them every opportunity to say they're going to try it. So, I would never let the history of non-adherence stop me from providing it unless they actively told me.* [Participant #9]

Performance Measurement With Audit and Feedback

Some Providers Prefer Dichotomous Performance Measures, Whereas Others Prefer Performance Measures That Incorporate Risk Categorization of Patient Groups

The team proposed a new performance measure consistent with the new guidelines that would provide weighted performance assessment. In this system, patients for whom statin treatment

was particularly likely to prevent a heart attack or stroke would be given more credit in evaluation. Providers had mixed feelings about the proposed VA performance measure, particularly the idea of weighting the performance measure to reflect risk categorization of patient groups (based on patients' ASCVD risk):

> *So, I could tell you with the measures, I will be honest with you; I don't like the idea of weighted. I like either you made it or you didn't…and I think either you're treating them appropriately or you're not…* [Participant #4]

Other providers preferred the proposed VA performance measure and that having a measure that "reflects" that difference may improve care:

> *...overall risk for some of these patients is higher or lower depending on which of these [risk] categories they fall into.* [Participant #7]

Providers Want More Credit for Shared Decision Making

Several providers were concerned with the lack of credit given for shared decision making in the proposed and existing VA performance measures. Most providers agreed that the high-risk patients, or the "no-brainers" as one provider put it, take less effort and time to convince to initiate and adhere to treatment because their risk is more palpable. Rather, it is the patients who:

> *feel fine and they haven't had any negative outcomes yet [sometimes] are the hardest ones to get to comply and understand, educate about what's in their best interest...* [Participant #9]

One provider specifically made the connection between the way pay-for-performance structures are designed and the lack of consideration given to shared decision making:

> *[Patients that fall in the intermediate risk category] You use a lot of energy with and you're really not capturing that much value from the standpoint of, whatever it's going to be, an A or money or whatever it is at the end that you get as your carrot. I don't know how you would do it any other way that I think makes sense. I don't think most of us are in it for the A or the money.* [Participant #12]

Providers Feel That Hierarchical Patient-Level Feedback Is Most Useful Within Audit and Feedback Reports

Providers regularly receive audit and feedback of their care within this clinic, usually in the form of printouts of tables of care provided. We attempted to understand how the new guidelines might alter the best way to provide audit and feedback. Providers generally preferred the audit and feedback report when broken down into component parts, indicating first, how the provider fares on each individual goal (ie, the percentage of the provider's patients with ASCVD that are on a moderate or high-dose statin) and second, broken down by patient fallout, with the highest risk patients listed first, and the lower-risk patients listed last. Providers indicated that listing out patients that did not meet the guideline by risk category would be more actionable than having a single list of patients not meeting guidelines, as members of the PACT team would then be better able to triage follow-up phone calls. As one provider explained:

> *It does help you gauge again from the standpoint of, where do you least want to make mistakes, with the people that have significant disease already and if you had someone with very, very low risk taking a statin, it's not going to be the worst thing in the world. I mean, you're not happy about it, but I think that is important to see the breakdown.* [Participant #9]

Several individuals indicated that comparing providers by their percentile of measures met is not motivating, in part, because it can be difficult to distinguish who they are being compared with, whether it is providers at the local level, or providers at the system level. Another provider mentioned that delayed receipt of the report also decreases impact on provider behavior, stating:

> *I think there is a big enough disconnect between the guidelines and the results coming out of it.* [Participant #12]

Providers Would Value Audit and Feedback More if It Were Used to Help Their Care More Directly

Participants generally wanted performance measures and audit and feedback reports to be more tightly linked to ways to help the providers improve practice in response:

> *I want the researchers or whoever's pulling this out for me, if somebody's in the highest percent I want you to interview them and tell me what...are they doing to be in that percentile. I'm not kidding you...Clearly somebody's in the top percentile. What are they doing?...It's like, don't just tell me where I'm at, tell me how to be better and do that by using this to find out who's doing better...* [Participant #4]

Providers wished there would be a greater commitment toward teaching them about new guidelines and changes in practice, particularly those moving toward incorporating patient risk and shared decision making. On the basis of interviews, we found that providers are willing to adopt risk-based clinical guidelines and accompanying components if they are designed with care and are presented to providers in a clear and useful manner:

> *It's more than just flipping on a switch and having some PowerPoint slides. I think that you really need to help clinicians move towards that, help them understand it, give them some strategies, give them some confidence for how to move in that direction because...it's another paradigm shift that we need to be making, but I think we need help in order to get there.* [Participant #2]

Discussion

Principal Findings

We developed a system to create a multicomponent implementation intervention that was to be user-centered and evidence based. Our system helped us identify ways to improve aspects of the intervention and develop new ones (see Table 2).

Textbox 1. Design considerations for future interventions.

Performance improvement suggestions

- **Representative quote:** "*So, don't just tell me these are wrong, tell me what I need to do, give me useful information so I now know, 'Oh, I need to call that patient and double it,' versus whatever, I mean, just to say they're on 40, that doesn't mean anything to me. So, that's the key thing, to make this helpful tell me what I need to do, because honestly, the more I know what I need to do, the more I can hand this to my nurse and say, 'Hey! Call that patient, order a blood draw, tell him what it is, I'm going to change their med,' make it so I don't have to do anything, yeah, there we go.*" [Participant #4]
- **Proposed response:** Include specific and actionable performance improvement suggestions

Provider comparison group

- **Representative quote:** "*It tells me my provider percentile; I can never tell if it's VA or local, okay so that's part of it, so that's one thing I would want to know VA or local.*" [Participant #12]
- **Proposed response:** Clarify provider comparison group

Patient fallout organization

- **Representative quote:** "*So, if you are going to work through a list, you want to start from the top and work your way down kind of deal. So, I think that that would be helpful because you're, at least at the outset, you're going to identify the most important areas for intervention…So, I do think that's helpful from just like a time management perspective, like start here and then over the next six months we'll get through everybody, but at least we'll start at the top and work down to the people who are maybe less of a priority as far as you know statin and cardiovascular risk reduction.*" [Participant #7]
- **Proposed response:** Organize patient fallout by risk category

Accounting for complicated patients

- **Representative quote:** "*It depends if we give credit for having...if we could include documented adverse drug reaction to giving you credit, then that would be good or just taking those people out altogether, you know, so they're not even in the, they're not even in the denominator, um because you know, there are a lot of people who have statin, and this is where it's provider, you know, it is provider. If you don't ask and you don't know, and you just keep pounding someone with statin and they're feeling miserable, it's not the right thing to do. So, if you're not aware of the potential side effects or you're not asking and you're not dealing with it, then your numbers may look better but you may not be doing the patient a service. So, I would say if the goal includes, if you get credit for at least a documented adverse drug reaction, then I'd be fine with those numbers. If not, they need to come out of the denominator, if not, the goal needs to be a little bit lower or I would recommend it be lower.*" [Participant #9]
- **Proposed response:** Think of mechanism to account for complicated patients in performance measure and subsequently in the audit and feedback report

In short, we found that providers were interested in changing their care but needed support in doing so.

Our team incorporated feasible design suggestions into the prototype intervention, particularly when there was general agreement among providers on a given design element and it aligned with design and user experience best practices. Consistent with previous research [22,23], providers overwhelmingly preferred simple information, clear and accurate decision prompts, and logical ordering of information that aligned with their values and needs, such as including highest risk patients first on audit and feedback report fallout lists. More specifically, providers wanted to be able to accurately and rapidly use clinical decision support tools during the patient encounter without any loss to their autonomy [24]. Many of these wording or formatting suggestions were addressed in the second iteration of the clinical decision support tool and audit and feedback report template.

Some providers found the shift to new guidelines difficult, even when the guidelines were more closely aligned to their clinical perspective. For example, providers also felt guidelines don't recognize the most difficult aspects of their work, particularly the time and resource demands of shared decision making and introducing the concept of risk, which is strongly emphasized in the new guidelines [25,26]. In addition, providers requested more evidence, education, and resources to make any clinical change. Educational and training resources for both providers and patients were thought to be essential in effective shared decision making and, as a result, adherence to statin guidelines. In response, we implemented an educational seminar during a primary care meeting whereby differences between the guidelines were highlighted by way of a pocket guide [27] and explained in detail before the commencement of the intervention phase of the project.

Our work adds to, but is supported by, existing research in implementation science on guideline implementation and how to change clinician habits. Our findings align well with our underlying framework, the Clinical Practice Guidelines Framework for Improvement [12]. As that framework and other research suggests, we found barriers and wide variation in providers' knowledge, attitudes, and behaviors about the new guidelines [12,28]. Previous work has also found that providers find guidelines and performance measures demotivating, especially when they are not user-centered or well-aligned with the providers' goals of care [29,30]. Similarly, decision support tools regularly impact patient care but details of usability also have large effects on provider satisfaction and uptake [31,32]. Our work is one of a few studies that have attempted to

synthesize these diverse fields of research into a single intervention. Our findings were also unusual in noting the central divide between providers' desire for new guidelines for support and efficiency versus a sense that they are intended to remove providers' autonomy.

Limitations and Future Research

We sampled a small number of providers from one VAMC. Nonetheless, the providers who participated in the interviews for this project provided important insights that influenced both the type and content of the intervention later executed at this site. We expect our research design to be transferable to other sites, as user-centered design and qualitative methods both emphasize local context.

We were also limited in our ability to incorporate many of our findings into the intervention. At times, providers' opinions and preferences were at odds. Thus, our team needed to prioritize and rank feedback, accommodating feasible design suggestions with strong provider consensus, and vetoing design elements that were too provider-specific, acknowledging that a provider-specific interface is not feasible within the health system. Relatedly, there were requests for user-friendly features that were technologically infeasible. Consequently, we have identified future design considerations for each of the above domains (see Textbox 1) that were outside the scope of this project but could be considered in other projects.

Finally, the purpose of this study was to follow a user-centered design approach to capture the needs and preferences of providers in the final intervention design. Though beyond the scope of this study, future research should examine the effectiveness of similar multicomponent implementation interventions.

Conclusions

The guideline implementation planning process provided important insights about the refinement of the intervention plan. By combining qualitative methods and user-centered design principles, we could understand the needs and preferences of providers and modify prototypes to increase their acceptability and usability in practice. Our findings allowed us to target several factors providers reported as being important determinants to the uptake of and adherence to clinical practice guidelines. The qualitative process of working with providers also allowed us to identify future design considerations for multicomponent guideline implementation interventions.

Acknowledgments

This work was supported through funding awarded to JBS by the VA Quality Enhancement Research (QUERI) Program (Grant #QUE 15-286) and VA Health Services Research and Development (HSR&D; Grants # IIR 15-432; CDA 13-021). The authors would like to thank the clinical leaders, physicians, and pharmacists who participated in this study. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Department of VA.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Single-page VA/DoD and ACC/AHA Clinical Practice Guidelines educational tool.

[PDF File (Adobe PDF File), 25KB - humanfactors_v5i2e19_app1.pdf]

Multimedia Appendix 2

Clinical decision support tool.

[PDF File (Adobe PDF File), 157KB - humanfactors_v5i2e19_app2.pdf]

Multimedia Appendix 3

VA proposed performance measure.

[PDF File (Adobe PDF File), 21KB - humanfactors_v5i2e19_app3.pdf]

Multimedia Appendix 4

Audit and feedback template.

[PDF File (Adobe PDF File), 31KB - humanfactors_v5i2e19_app4.pdf]

Multimedia Appendix 5

User-centered design changes to intervention components.

[PDF File (Adobe PDF File), 24KB - humanfactors_v5i2e19_app5.pdf]

References

1. Centers for Disease Control and Prevention. Therapeutic Drug Use URL: https://www.cdc.gov/nchs/fastats/drug-use-therapeutic.htm [WebCite Cache ID 6qa36Otk0]
2. Goff Jr DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Jun 24;129(25 Suppl 2):S49-S73 [FREE Full text] [doi: 10.1161/01.cir.0000437741.48606.98] [Medline: 24222018]
3. Downs JR, O'Malley PG. Management of dyslipidemia for cardiovascular disease risk reduction: synopsis of the 2014 U.S. Department of Veterans Affairs and U.S. Department of Defense clinical practice guideline. Ann Intern Med 2015 Aug 18;163(4):291-297. [doi: 10.7326/M15-0840] [Medline: 26099117]
4. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling Jr JW, et al. Statin use for the primary prevention of cardiovascular disease in adults: US Preventive Services Task Force Recommendation Statement. J Am Med Assoc 2016 Nov 15;316(19):1997-2007. [doi: 10.1001/jama.2016.15450] [Medline: 27838723]
5. Hayward RA, Krumholz HM, Zulman DM, Timbie JW, Vijan S. Optimizing statin treatment for primary prevention of coronary artery disease. Ann Intern Med 2010 Jan 19;152(2):69-77. [doi: 10.7326/0003-4819-152-2-201001190-00004] [Medline: 20083825]
6. Sussman J, Vijan S, Hayward R. Using benefit-based tailored treatment to improve the use of antihypertensive medications. Circulation 2013 Nov 19;128(21):2309-2317 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.002290] [Medline: 24190955]
7. Hayward RA, Hofer TP, Vijan S. Narrative review: lack of evidence for recommended low-density lipoprotein treatment targets: a solvable problem. Ann Intern Med 2006 Oct 03;145(7):520. [doi: 10.7326/0003-4819-145-7-200610030-00010]
8. Krumholz HM, Hayward RA. Shifting views on lipid lowering therapy. Br Med J 2010;341:c3531 [FREE Full text] [Medline: 20667950]
9. Sussman JB, Vijan S, Choi H, Hayward RA. Individual and population benefits of daily aspirin therapy: a proposal for personalizing national guidelines. Circ Cardiovasc Qual Outcomes 2011 May;4(3):268-275. [Medline: 21487091]
10. Hendrani AD, Adesiyun T, Quispe R, Jones SR, Stone NJ, Blumenthal RS, et al. Dyslipidemia management in primary prevention of cardiovascular disease: current guidelines and strategies. World J Cardiol 2016 Feb 26;8(2):201-210 [FREE Full text] [doi: 10.4330/wjc.v8.i2.201] [Medline: 26981215]
11. Kortteisto T, Kaila M, Komulainen J, Mäntyranta T, Rissanen P. Healthcare professionals' intentions to use clinical guidelines: a survey using the theory of planned behaviour. Implement Sci 2010;5:51. [Medline: 20587021]
12. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PC, et al. Why don't physicians follow clinical practice guidelines? J Am Med Assoc 1999 Oct 20;282(15):1458. [doi: 10.1001/jama.282.15.1458]
13. Ritter FE, Baxter GD, Churchill EF. Foundations for Designing User-Centered Systems: What System Designers Need to Know About People. London: Springer; 2014.
14. Francke AL, Smit MC, de Veer AJ, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. BMC Med Inform Decis Mak 2008;8:38. [doi: 10.1186/1472-6947-8-38]
15. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2015 Nov 27;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
16. Morse JM. The significance of saturation. Qual Health Res 1995;5(2):147-149. [doi: 10.1177/104973239500500201]
17. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
18. Veterans Health Administration. VHA Operations Activities That May Constitute Research. Research.va.gov 2018 [FREE Full text]
19. Longo L, Kane B. A novel methodology for evaluating user interfaces in health care. 2011 Presented at: Proceedings of the 24th IEEE International Symposium on Computer-Based Medical Systems; JUNE 27-30, 2011; Bristol, United Kingdom. [doi: 10.1109/CBMS.2011.5999024]
20. Harrison MB, Légaré F, Graham ID, Fervers B. Adapting clinical practice guidelines to local context and assessing barriers to their use. CMAJ 2010 Feb 09;182(2):E78-E84 [FREE Full text] [doi: 10.1503/cmaj.081232] [Medline: 19969563]
21. List D. The consensus group technique in social research. Field Methods 2016 Jul 24;13(3):277-290. [doi: 10.1177/1525822X0101300304]
22. 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]
23. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implement Sci 2014 Jan 17;9:14 [FREE Full text] [doi: 10.1186/1748-5908-9-14] [Medline: 24438584]

24. Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decis Support Syst 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
25. Xu Y, Wells PS. Getting (along) with the guidelines: reconciling patient autonomy and quality improvement through shared decision making. Acad Med 2016 Jul;91(7):925-929 [FREE Full text] [doi: 10.1097/ACM.0000000000001101] [Medline: 26839943]
26. Hess EP, Coylewright M, Frosch DL, Shah ND. Implementation of shared decision making in cardiovascular care: past, present, and future. Circ Cardiovasc Qual Outcomes 2014 Sep;7(5):797-803 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000351] [Medline: 25052074]
27. Lewin Group. 2014. VA/DoD clinical practice guideline for the management of dyslipidemia for cardiovascular risk reduction URL: https://www.healthquality.va.gov/guidelines/CD/lipids/VADoDDyslipidemiaCPG.pdf [WebCite Cache ID 6qa3iSWZj]
28. Gagliardi AR, Brouwers MC, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. Implement Sci 2011 Mar 22;6:26 [FREE Full text] [doi: 10.1186/1748-5908-6-26] [Medline: 21426574]
29. Damschroder LJ, Robinson CH, Francis J, Bentley DR, Krein SL, Rosland A, et al. Effects of performance measure implementation on clinical manager and provider motivation. J Gen Intern Med 2014 Dec;29 Suppl 4:877-884 [FREE Full text] [doi: 10.1007/s11606-014-3020-9] [Medline: 25234554]
30. Meyer GS, Nelson EC, Pryor DB, James B, Swensen SJ, Kaplan GS, et al. More quality measures versus measuring what matters: a call for balance and parsimony. BMJ Qual Saf 2012 Nov;21(11):964-968 [FREE Full text] [doi: 10.1136/bmjqs-2012-001081] [Medline: 22893696]
31. Wright A, Sittig DF, Ash JS, Erickson JL, Hickman TT, Paterno M, et al. Lessons learned from implementing service-oriented clinical decision support at four sites: a qualitative study. Int J Med Inform 2015 Nov;84(11):901-911. [doi: 10.1016/j.ijmedinf.2015.08.008]
32. Ash JS, Sittig DF, Guappone KP, Dykstra RH, Richardson J, Wright A, et al. Recommended practices for computerized clinical decision support and knowledge management in community settings: A qualitative study. BMC Med Infor Decis Mak 2012;12(1):6. [Medline: 22333210]

Abbreviations

ACC: American College of Cardiology
AHA: American Heart Association
DoD: Department of Defense
PACT: Patient Aligned Care Team
VA: Department of Veterans Affairs
VAMC: Veterans Affairs Medical Center
VHA: Veterans Health Administration

Edited by G Eysenbach; submitted 25.09.17; peer-reviewed by P Naidoo, H Lam; comments to author 09.12.17; revised version received 31.01.18; accepted 16.02.18; published 24.04.18.

Please cite as:
DeJonckheere M, Robinson CH, Evans L, Lowery J, Youles B, Tremblay A, Kelley C, Sussman JB
Designing for Clinical Change: Creating an Intervention to Implement New Statin Guidelines in a Primary Care Clinic
JMIR Hum Factors 2018;5(2):e19
URL: http://humanfactors.jmir.org/2018/2/e19/
doi:10.2196/humanfactors.9030
PMID:29691206

Viewpoint

Value of Face-to-Face Interactions Between Clinician-Educators and Patients or Students to Improve Health Care Education

Manisha Singh[1], MD

Division of Nephrology, Department of Internal Medicine, University of Arkansas For Medical Sciences, Little Rock, AR, United States

Corresponding Author:
Manisha Singh, MD
Division of Nephrology
Department of Internal Medicine
University of Arkansas For Medical Sciences
4301 West Markham St #501
Little Rock, AR, 72205
United States
Phone: 1 5012405804
Email: msingh@uams.edu

Abstract

The power and outreach of the media is enormous and has restructured our society today; the author acknowledges the impact and appreciates the outreach. However, I question the relative lack of focus on physical human interactions and express concern over future training efforts. I have compared and attempted to highlight the components of two interaction scenarios: those of teacher-student, and those of physician-patient. The physician-educators need to generate a discussion regarding the value of each interaction. As a teacher, there is value in online classrooms, and a different value in face-to-face interactions. Similarly, a physician can have major outreach impact by online tele-medicine and tele-education efforts, but in some instances, may need to have the human, physical interaction with the patient. The value of these interactions depends on the roles in which these interactions are experienced. Medical education training must incorporate an understanding of the unique value of different interactions.

(JMIR Hum Factors 2018;5(2):e15) doi:10.2196/humanfactors.9859

KEYWORDS

medical education; human factor; value in human interactions

Observations

The “human factor” is the essence of teaching and learning as well as that of clinical care. With the advent of media, we need to examine its role as it is taught and learnt, expressed as “face-to-face” versus “using technology” as a medium of interaction. The value of the medium used for an interaction has to be assessed based on the outcome expected from that interaction. In some circumstances, the use of media is very helpful. Conversely, sometimes media takes away value from the purpose of the interaction.

Questions

Experience of classroom teaching

I arrive early, formally dressed, ready with a PowerPoint presentation. The PowerPoint consists of information most relevant for exams, with example questions and case-based presentations. I walk into a class of 10 students and teach from the podium. There is a live feed on. The other 155 students are at different places, logging in from their places of choice. Many review the lecture videos later. After class, they critique my lecture for relevance to the board preparation. My future presentations will be based on the critiques from these students. The same topics have been covered well by leading experts in the area with freely available videos. Between the textbooks and internet resources, where is the need of an educator standing in class on a particularly rainy morning?

Is the experience of the students in class different from those online?

My experience in the clinic is also similarly thought provoking with the advent of media.

Patients in a subspecialty clinic

The patient can be treated without meeting the doctor in some cases. For instance, for a patient with chronic kidney disease we can order laboratory tests and the primary care physicians’

note can tell us about the exam findings needed to prescribe appropriate medications. For routine subspecialty assessments we may not need a physical exam. The vitals can be texted in. An effective, guideline-based management plan can be put in place without any person-to-person interactions. The patient can be emailed the medication side effect list and get an online consent, which may be a safe process as it related to liability.

The subspecialty can do away with most barriers that lead to conflicts. The biases of nationality, gender, and language can all be transcribed over. The human factor can thus be eliminated entirely in both of these scenarios. However, there is some value in each interaction: using media or not using media; with or without the human factor.

The Human Factor

We, as people, have an identity of self and also an identity in the roles we play. Each role comes with its society-approved behavior patterns: roles of spouses, children, or being parents; roles of identifying with a particular gender, race, or nationality. Each role comes with its own set of expected behaviors, attire, and stage setting. The role of a physician comes with props like the white coat and the stethoscope, the stage setting of a clinic or hospital, and in recent advances, tele-health support. The role of a student may come with a recognition of some degree of need for training and accepting a person as a teacher (in the setting of a classroom or through online presence).

My pedagogy mentors emphasized a balance of head, hands, and heart when I started my career in clinical education. While the head and hands are approachable through media, we may need to look closer at the heart for training and treatment purposes. Regardless of the modality of interaction (face-to-face or via media), the perceived roles of human interactions can vary between those of student-teacher, physician-patient, and vendor-consumer.

Each role has a set of rights and duties assigned, and also some combination of the following individual components: required knowledge (head), the required skills (hands), and the required attitudes that go with this (heart). An effective combination of these factors is what constitutes the human factor: the person you meet.

Student-Teacher Scenario

The teacher is more than a container/dispenser of information. The personality of an educator is unique. The scales we use to assess the caliber of a teacher uses only objective data: how well the students do on exams and their assessment of the course content with respect to national exams. The subjective experience of being a student in classroom, facing a teacher, and the outcome of this interaction is a human relationship; its value is very hard to express/assess. The student learns more than the subject. The teacher, in turn, gets value in that experience. When a teacher teaches in class, he/she brings with that PowerPoint presentation, a human factor which responds and relates to the student.

Regarding the content of education material, recent advances with tele-education have expanded educational outreach beyond imagination [1]. Initially we could have critiqued that the content available is not peer reviewed and lacks the authenticity of experts opining on the content [2]. This line of thinking has changed completely with online discussions with experts on various platforms, as seen with online journal clubs, media groups, and scientific communities. There is an admirable turnaround time of a scientist's question being answered in the digital world.

Physician-Patient Scenario

The physician that is present in the clinic has a relationship with a patient he/she faces in a patient-doctor relationship, which is different from communicating with the same patient through media. Although tele-education and tele-medicine are enormously helpful in many cases, there is something to be said for the personal interaction.

What is the Value of Holding a Hand, Making Eye Contact, and Talking to Another Person?

The physician, much like the teacher, aspires to "make the patient feel better.'' There is a sense of aspiring for a higher purpose in life. Most physicians in the role of "the doctor" care deeply about the best state of health possible for the patient in accordance with the patient's own challenges. To achieve this, they train to the extent of their abilities and challenge limitations. Their interactions are based on the roles perceived by each. A physician can see himself in the role of a caregiver, or that of a vendor, regardless which form of communication he/she chooses to use. A patient can see him/herself as a person needing comfort or a consumer, again, regardless of how he/she is communicating with the physician. These perceived roles can be seen similarly in any situation, be it personal interaction or through a digital presence. The human factor introduces a relationship between roles. We need to understand the roles of the patient-doctor versus roles of the vendor-consumer. In any scenario, either relationship may be perceived.

If the physician is seen in the role of a vendor, the patient assumes the role of a consumer, and must look for a certain specific objective value in the encounter [3]. A vendor's motivation is different than that of a physician. In this scenario, there is not much value in the time spent sitting silently holding a patient's hand after delivering bad news. A 15-minute encounter, encompassing a full review of systems, an exam, review of lab results, and assessment and management plan, has no financial incentive to add to the question, "How do you feel today?" The best gain scenario is to identify a problem list quickly and focus on the exam and a treatment plan, minus the human-factor.

The patient, in the role of a patient, is vulnerable and expects to be comforted. Armed with internet knowledge, the patient may require more detailed discussions than previously anticipated [4]. In the role of the consumer, the patient is able to critique the physician for "professionalism" which is a very subjective interpretation. For the "consumer" the option of using media may present the best-case scenario, while the role of a "patient" may need a human interaction with a person in the physical role of a doctor.

Unifying Concepts

The role of basic courtesy is crucial in either case, but is there any added value to physically meet a teacher/physician? Is an email the same thing as meeting someone in person? How does Skype compare to face-to-face meetings? Are we gaining in economics and losing in humanities?

Will Either Situation Add Value to The Objective of That Interaction?

Where there is no health care, or in areas where it is difficult to obtain subspecialty quality care, tele-medicine certainly is the most valuable tool. Similarly, for lectures, online tools are very valuable for distant outreach. However, there is still a place for training in human interactions for many scenarios in which the value is in the absolute human interaction.

There are places in which this is helpful, and there may be places that this is time consuming, redundant, and expensive. In changing times, with the power and outreach of media, a discussion that fully incorporates the advantages of media and the possibilities of including the human factor for an effective classroom and clinical experience is warranted.

Acknowledgments

I would like to thank Mr Yashwant Singh and Shree Sharma, MD, for support with the writing and thought process refinement.

Conflicts of Interest

None declared.

References

1. Topf JM, Hiremath S. Social media, medicine and the modern journal club. Int Rev Psychiatry 2015 Apr;27(2):147-154. [doi: 10.3109/09540261.2014.998991] [Medline: 25906989]
2. Allahwala U, Nadkarni A, Sebaratnam DF. Wikipedia use amongst medical students - new insights into the digital revolution. Med Teach 2013 Apr;35(4):337. [doi: 10.3109/0142159X.2012.737064] [Medline: 23137251]
3. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
4. Jacob J. Consumer access to health care information: its effect on the physician-patient relationship. Alaska Med 2002;44(4):75-82. [Medline: 12650084]

Edited by P Santana-Mancilla; submitted 16.01.18; peer-reviewed by CS Hung, S Hiremath; comments to author 01.02.18; revised version received 19.02.18; accepted 26.02.18; published 20.04.18.

Please cite as:
Singh M
Value of Face-to-Face Interactions Between Clinician-Educators and Patients or Students to Improve Health Care Education
JMIR Hum Factors 2018;5(2):e15
URL: http://humanfactors.jmir.org/2018/2/e15/
doi:10.2196/humanfactors.9859
PMID:29678803

Original Paper

Translation, Cross-Cultural Adaptation, and Validation of the Malay Version of the System Usability Scale Questionnaire for the Assessment of Mobile Apps

Muhamad Fadhil Mohamad Marzuki[1], MBBS, MPH; Nor Azwany Yaacob[1], MD, MCM; Najib Majdi Yaacob[2], MD, MPH, Dr PH

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

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

Corresponding Author:
Muhamad Fadhil Mohamad Marzuki, MBBS, MPH
Department of Community Medicine
School of Medical Sciences
Universiti Sains Malaysia
Kubang Kerian
Kelantan, 16150
Malaysia
Phone: 60 97676621
Fax: 60 7676654
Email: fadhilmarzuki@gmail.com

Abstract

Background: A mobile app is a programmed system designed to be used by a target user on a mobile device. The usability of such a system refers not only to the extent to which product can be used to achieve the task that it was designed for, but also its effectiveness and efficiency, as well as user satisfaction. The System Usability Scale is one of the most commonly used questionnaires used to assess the usability of a system. The original 10-item version of System Usability Scale was developed in English and thus needs to be adapted into local languages to assess the usability of a mobile apps developed in other languages.

Objective: The aim of this study is to translate and validate (with cross-cultural adaptation) the English System Usability Scale questionnaire into Malay, the main language spoken in Malaysia. The development of a translated version will allow the usability of mobile apps to be assessed in Malay.

Methods: Forward and backward translation of the questionnaire was conducted by groups of Malay native speakers who spoke English as their second language. The final version was obtained after reconciliation and cross-cultural adaptation. The content of the Malay System Usability Scale questionnaire for mobile apps was validated by 10 experts in mobile app development. The efficacy of the questionnaire was further probed by testing the face validity on 10 mobile phone users, followed by reliability testing involving 54 mobile phone users.

Results: The content validity index was determined to be 0.91, indicating good relevancy of the 10 items used to assess the usability of a mobile app. Calculation of the face validity index resulted in a value of 0.94, therefore indicating that the questionnaire was easily understood by the users. Reliability testing showed a Cronbach alpha value of .85 (95% CI 0.79-0.91) indicating that the translated System Usability Scale questionnaire is a reliable tool for the assessment of usability of a mobile app.

Conclusions: The Malay System Usability Scale questionnaire is a valid and reliable tool to assess the usability of mobile app in Malaysia.

(JMIR Hum Factors 2018;5(2):e10308) doi:10.2196/10308

KEYWORDS

usability; System Usability Scale; Malay; questionnaire translation; questionnaire validation; mobile app

Introduction

The advancement of communication technologies has changed the way people search for and find information. This is especially prevalent in the case of health-related information. Consequently, health providers should update their health education and promotion strategies to disseminate information from conventional printed material such as pamphlets and flip charts, to more interactive and updated material such as mobile apps [1]. Mobile apps have the advantage of being widely available soon after development through multiple platforms. The usability of the mobile app in question plays an important role in determining its effectiveness to improve health knowledge and awareness. An app must not only be user-friendly, but it should also attract users.

Usability is defined as the extent to which a product can be used by specified users to achieve specific goals effectively and efficiently as well as providing user satisfaction in a specified context of use [2]. Questionnaire surveys are among the established and acceptable methods for system usability evaluation [3]. Usability consists of 5 quality attributes of the system which assess how easy user interfaces are to use [4], namely learnability, efficiency, memorability, system errors, and user satisfaction.

Generally, there are two methods of assessing the usability of a product, expert reviews and usability testing [5]. Many questionnaires have been developed for usability assessment of computer-based interfaces, websites, apps, or any software or hardware with which users interact. These include the After Scenario Questionnaire (ASQ), Computer System Usability Questionnaire (CSUQ), and the Usefulness, Satisfaction and Ease of Use (USE) questionnaire [6]. The usability questionnaires recommended for the assessment of mobile apps can range from two simple post-test questions, to standard questionnaires such as the Post-Study Usability Questionnaire (PSSUQ) or the System Usability Scale (SUS) [7,8].

The System Usability Scale (SUS) is one the most widely used questionnaires to assess the usability of a system or product [9]. It was developed by John Broke in 1986 in response to the demand of many industries for a simple, quick, and cost-effective method to assess the usability of a system [10]. It has been utilized in various surveys to determine the usability of wide range of user interfaces such as standard operating system-based software interfaces, Web-pages, mobile apps, and networking equipment [6]. Originally, the SUS was developed for Digital Equipment Co Ltd customers who are the native English speakers [9]. The SUS questionnaire has since been translated into many languages including Spanish, French, Dutch, Portuguese, Slovenian, Persian, German, and more recently Indonesian. All translated versions have shown similar internal reliability to the original English version [11].

To the best of the authors' knowledge, there are no studies reporting the translation of the SUS questionnaire into Malay, despite the widespread usage of this questionnaire across the world. It is crucial to have a SUS questionnaire in the local language to accurately capture the thoughts, feelings, perceptions, behaviors, and attitudes of local users towards the usability of the tested product. Different cultures can interpret similar words or phrases in a different manner, therefore the translation used in this study takes into consideration the linguistics of the questionnaire, as well as the cross-cultural adaptation needed to maintain the validity of the questionnaire [11]. Thus, the objective of this study is to translate and validate the original English version of the SUS into Malay.

Methods

Overview

The SUS was developed by John Brooke in 1986 [10] and consists of a 10-item questionnaire scored on a 5-point Likert scale from 0 (strongly disagree) to 5 (strongly agree). The questionnaire is arranged to alternate between positive and negative statements to avoid habitual bias from the respondent. The score contribution for the odd items (the positive statements) is the scale position minus 1 and the contribution for the even items (the negative statements) is 5 minus the scale position. The overall score is calculated from a sum of all item scores multiplied by 2.5 and can range from 0 to 100. A system or product that received score of 68 and above is considered to have good usability [10].

Adaptation Process

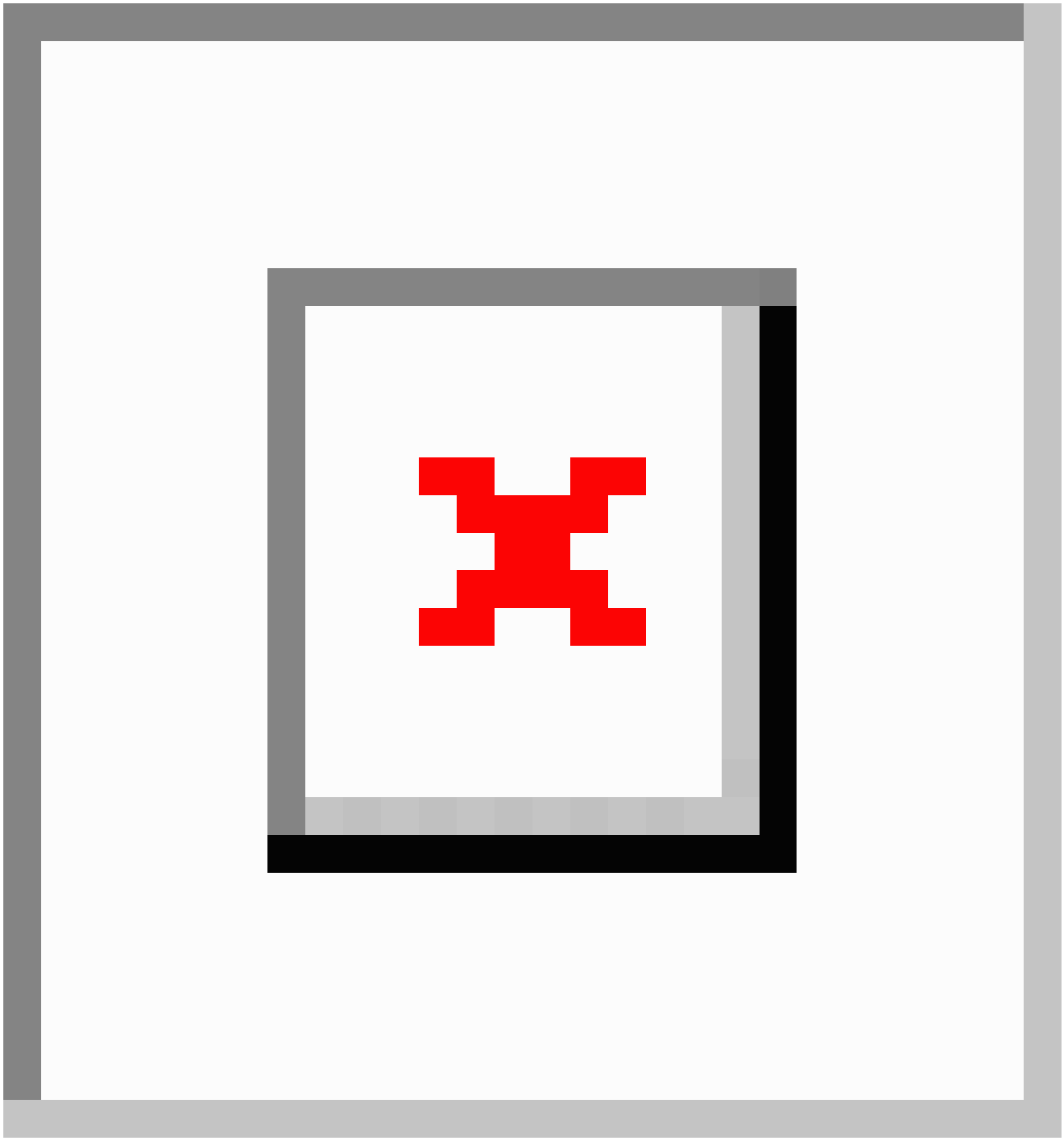
The original SUS questionnaire was translated into Malay using international guidelines for cross-cultural adaptation to ensure the quality of the translated version and its consistency of meaning to the original version [12]. First, the forward translation process (from English to Malay) was conducted by two translators and a report of the translation was produced by both translators. The two translations were synthesized into one document after a thorough discussion which addressed any gaps or differences between the two reports.

The original and translated versions of the SUS questionnaire were given to two groups of native Malay speakers who spoke English as their second language. Each group consisted of 8 translators who received either the original or translated questionnaire version and then performed either the forward or backward translation respectively. The forward and backward translation discrepancies were reconciled, and cross-cultural adaptation was done to derive the final version. Since the purpose of translating the SUS questionnaire is to assess the usability of mobile apps, the word "system" has been changed to "mobile application" in the survey. The Malay term for this is "aplikasi mudah alih," hence the adapted questionnaire is called *Skala Kebolehgunaan Aplikasi Mudah Alih* (SKAMA) in Malay.

Validation Process

The SKAMA questionnaire was subsequently validated in terms of its content validity, face validity, and reliability (internal consistency). Content validation aims to assess the relevancy and representativeness of each item to a specific domain by a panel of experts. In this context, it will assess the relevance of all 10 items in the SKAMA to represent the usability domain.

Figure 1. Flowchart of the validation process.



Content validation of the SKAMA questionnaire was conducted by 10 experts (including 2 mobile app developers) who were asked to give a score of 1 (item not relevant) to 4 (item very relevant), based on the relevancy of the translated items in the SKAMA, to assess the usability of a mobile app. Scores of 3 and 4 were recategorized as 1 (relevant) and scores of 1 and 2 as 0 (not relevant). The content validity index (CVI) was computed by calculating the scale average [13]. Figure 1 gives an overview of the validation process.

Face validation testing, which aims to assess the clarity and comprehensibility of the translated items, was conducted by 10 target users. The users were asked to give score from 1 (item not clear and not understandable) to 4 (item very clear and understandable) based on the clarity and comprehensibility of the translated items in the SKAMA questionnaire. Scores of 3 and 4 were recategorized as 1 (clear and understandable) and scores of 1 and 2 as 0 (not clear and understandable). The face validity index (FVI) was computed by calculating the scale average [13]. Reliability testing was conducted on 49 respondents based on a minimum sample size estimation to assess the internal consistency [14]. They were asked to use the SKAMA to assess the usability of the Facebook mobile app on their mobile phone. The reliability analysis was computed using R software. All three validation tests performed on the SKAMA questionnaire were conducted using an online Google Form where the link was sent to each respondent via a personal WhatsApp (for the validation test) or a group WhatsApp (for the reliability test) to facilitate the data collection.

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

Results

In the translation of the SUS questionnaire, the word "system" was changed to the Malay word for "mobile application," namely "*aplikasi mudah alih*," as the Malay adaptation of the SUS questionnaire is intended to determine the usability of mobile apps. The CVI (Table 1) and FVI (Table 2) of SKAMA were calculated to be 0.91 and 0.94 respectively. The CVI and FVI score of above 0.83 for both tests indicates that all items in the questionnaire are relevant to the domain, clear, and comprehensible for the target users [13,15].

The reliability testing was conducted using 53 target users (the minimum estimated sample size was 49 respondents) who responded to the online questionnaire via a URL link sent to them. The age of the respondents ranged from 23 to 60 years. The majority of the target users worked for the government and have a tertiary education. Table 3 shows the characteristics of the target users who responded to the online questionnaire.

Table 1. Content validity index based on the rating of the relevancy of items by 10 experts.

Item	E[a]1	E	E3	E4	E5	E6	E7	E8	E9	E 10	I-CVI[b]
Q1	4	4	4	3	4	4	4	4	3	4	1.00
Q2	2	3	4	3	4	3	4	4	4	1	0.80
Q3	4	4	2	4	4	4	4	4	4	4	0.90
Q4	4	4	4	4	3	4	4	4	4	1	0.90
Q5	3	4	3	4	4	4	4	1	4	4	0.90
Q6	4	4	4	4	4	4	4	4	4	4	1.00
Q7	4	3	4	4	3	4	3	3	4	4	1.00
Q8	4	1	3	4	3	3	4	4	1	3	0.80
Q9	3	3	3	4	3	4	4	4	4	4	1.00
Q10	4	1	4	4	3	2	4	3	4	3	0.80
Content validity index average											0.91

[a]E: Expert.

[b]i-CVI: Item Content Validity Index.

Table 2. Face validity index based on the rating of the clarity and comprehensibility of items by 10 target users.

Item	R[a]1	R2	R3	R4	R5	R6	R7	R8	R9	R 10	I-FVI[b]
Q1	3	4	4	4	4	4	4	2	4	4	0.90
Q2	3	4	4	3	4	4	4	3	4	4	1.00
Q3	4	4	4	4	4	4	4	3	4	4	1.00
Q4	4	4	4	4	4	4	4	3	4	4	1.00
Q5	3	4	3	4	4	4	4	1	4	4	0.90
Q6	2	3	4	4	3	4	4	1	3	3	0.80
Q7	3	3	4	3	4	4	4	2	3	4	0.90
Q8	4	4	4	4	3	4	4	2	4	4	0.90
Q9	3	4	4	4	4	4	4	3	4	4	1.00
Q10	3	3	4	3	4	4	4	3	4	4	1.00
Face validity index average											0.94

[a]R: Rater.

[b]I-FVI: Item Face Validity Index.

Table 3. Characteristics of target users (N=53).

Characteristic	Value
Age (years), mean (SD)	39.4 (10.46)
Highest education, n (%)	
Primary School	1 (1.9)
Secondary School	7 (13.2)
Tertiary education	45 (84.9)
Occupation, n (%)	
Government	38 (71.7)
Private	6 (11.3)
Pensioner	3 (5.7)
Unemployed	6 (11.3)

Table 4. The internal consistency of the total-item statistics.

Item	Scale mean if item deleted	Scale variance if item deleted	Corrected item total correlation	Cronbach alpha if item deleted
Q1	35.94	25.478	0.416	.85
Q2	35.70	26.830	0.460	.84
Q3	35.38	26.816	0.674	.83
Q4	35.51	27.370	0.459	.84
Q5	36.02	24.134	0.651	.82
Q6	36.40	24.205	0.653	.82
Q7	35.79	26.475	0.469	.84
Q8	35.79	25.245	0.637	.83
Q9	35.75	22.881	0.793	.81
Q10	36.19	24.887	0.429	.85

The Cronbach alpha for the SKAMA questionnaire was determined to be .85 (95% CI 0.79-0.91) which is similar to the original English SUS questionnaire [10]. A higher alpha value indicates a higher internal reliability of the questionnaire and value more than .70 is acceptable as satisfactory internal reliability [16]. The Cronbach alpha for the questionnaire if an item is deleted (from the questionnaire) also remains consistent without significant difference (Table 4) indicating good internal reliability of the developed questionnaire.

Discussion

The concept of system usability was first coined in the 1980s, in the field of human-computer interaction, when the first personal computer was developed [17]. Usability is the quality attributes of a system which assess how easy a system interface is to use [4]. These attributes include:

1. The learnability of the system (ie, how well users can learn and use a product to achieve the intended goals [18]).
2. The efficiency of the system (ie, how quickly users can perform the task once they learn the design).
3. The memorability of the system (ie, how easily the user can re-establish proficiency when they return to the system after a period of not using it).
4. The errors from using the system.
5. User satisfaction when using the system.

Ideally the usability evaluation of a system should be considered in every step of prototype development, a process which consists of iterative cycles of prototyping, design, and validation [19]. The usability of a developed system can be evaluated either by expert reviews or by usability testing [5]. Expert reviews can be conducted using heuristic checklists, cognitive walkthrough, and guidelines. This is dependent on the experts' knowledge and experience and therefore this may not reflect the users' perception of product usability. On the other hand, questionnaires are specifically developed to explore a construct that cannot be measured directly, such as attitude and practice, as well as the usability of a system. Creating a new questionnaire requires a concerted effort from team members, additional costs, and is time consuming. Therefore, researchers are recommended to adapt established, appropriate, and available questionnaires with documented validity in other languages. Literal translation, however, is not sufficient to produce an equivalent questionnaire. The questionnaire must have a good linguistic translation and must be adapted for cultural differences to maintain the content validity [11]. This is referred to as the cross-cultural adaptation of a questionnaire [20]. Validation, on the other hand, aims to ensure that the translated version

questionnaire has the same equivalent properties for measuring the construct as the original version. Cross-cultural adaptation to ensure the integrity of the questionnaire is retained, as translation can be problematic, especially when the two languages have nonequivalent words. It is especially important to take into account the fact that different cultures may interpret similar words or phrases in the questionnaire differently and therefore the intended meaning of the items in the questionnaire could be altered from the original version.

Malay is the native language in Malaysia, although multi-ethnic groups do exist. It is for this reason that this study aimed to translate the SUS questionnaire into Malay for use in Malaysia. The Malay version of the SUS, SKAMA, was reviewed by experts in the field, which included mobile app developers, as the aim is to use this translation for the assessment of mobile apps. Therefore, in the translation, the word "system" in the original SUS was replaced with the Malay term for "mobile application." The experts reviewed the SKAMA questionnaire content in relation to assessment of the usability of mobile apps, taking into account the considerations of local users. Face validity testing tested the clarity of the items to assess usability of mobile app from the target user point of view. Developers and experts in mobile apps may have a different view of system usability compared to the public users, who are the target users when new apps are developed. These two different groups of reviews help to ensure content coverage, while taking into consideration the comprehensibility of the items in the questionnaire to the target user. The high CVI and FVI of SKAMA thus indicates the content is well adopted into local context and translated using clear and understandable sentences.

The reliability of a questionnaire contributes to the validity of it and measures the stability of the questionnaire in terms of consistency of the response. Internal consistency is one of the reliability components used to measure the extent of which the items are measuring the same thing. The most common estimation of internal consistency is the Cronbach alpha coefficient [21]. The high Cronbach alpha value in this study indicates that SKAMA is a reliable tool to assess the usability of a mobile app. The consistent item statistics indicates that all 10 items are measuring a same domain, which is the usability of mobile app. Thus, the SKAMA questionnaire has equal reliability with similar Cronbach alpha values to the original SUS questionnaire and slightly higher values compared to the Indonesian version of SUS [9,11].

In conclusion, the SKAMA questionnaire is a valid tool to measure the usability of a mobile app for a Malay speaking population. SKAMA may also be used to assess other systems' usability by rephrasing the word "mobile application" back into "system" as in the original SUS.

Acknowledgments

We would like to express our gratitude to the lecturers and postgraduate candidates from the Department of Community Medicine, Unit of Medical Biostatistics and Research Methodology, and Department of Medical Education, Universiti Sains Malaysia, for the guidance, expert opinion, and contributions in the questionnaire adaptation process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Actual questionnaire Skala Kebolehgunaan Aplikasi Mudah Alih (SKAMA) or Mobile Application Usability Scale.

[PDF File (Adobe PDF File), 232KB - humanfactors_v5i2e10308_app1.pdf]

References

1. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995-9997. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
2. Usability Net. 1998 Jan 16. Guidance on Usability URL: http://www.usabilitynet.org/tools/r_international.htm [accessed 2018-04-30] [WebCite Cache ID 6z3uFSzUd]
3. Martins A, Rosa A, Queirós A, Silva A, Rocha N. European Portuguese Validation of the System Usability Scale (SUS). Procedia Computer Science 2015;67:293-300.
4. Nielsen J. Nielsen Norman Group. 2012 Jan 16. Usability 101: Introduction to Usability URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2018-04-30] [WebCite Cache ID 6z3uPGDYt]
5. Lang T, Zrobok M. PeakXD. 2016 Dec 7. Usability testing versus expert views: a comparison of usability evaluation methods URL: https://www.peakusability.com.au/articles/usability-testing-versus-expert-views [accessed 2018-04-30] [WebCite Cache ID 6z3uVUYhB]
6. Bangor A, Kortum P, Miller J. An Empirical Evaluation of the System Usability Scale. International Journal of Human 2008;24(6):574-594.
7. Mifsud J. Usabilitygeek. 2016 Mar 21. Usability Testing of Mobile Applications: A Step-By-Step Guide URL: https://usabilitygeek.com/usability-testing-mobile-applications/ [accessed 2018-04-30] [WebCite Cache ID 6z3uhlJbo]

8. Spool J. User Interface Engineering. 2006 Mar 23. Two Simple Post-Test Questions URL: https://www.uie.com/brainsparks/2006/03/23/two-simple-post-test-questions/ [accessed 2018-04-30] [WebCite Cache ID 6z3urP00n]
9. Brooke J. Journal of Usability Studies.: Journal of Usability Studies; 2013. SUS: A Retrospective URL: http://uxpajournal.org/sus-a-retrospective/ [accessed 2018-04-30] [WebCite Cache ID 6z3vROlWP]
10. Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry 1996;189(194):4-7.
11. Zahra S, Budi S. An Indonesian adaptation of the System Usability Scale (SUS). : IEEE; 2016 Presented at: International Conference in Advanced Computer Science and Information Systems (ICACSIS); 2016; Malang, Indonesia.
12. Sousa V, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. Journal of Evaluation in Clinical Practice 2011;17(2).
13. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
14. Wan Nor Arifin. 2017. Sample size calculator (Version 2) URL: http://wnarifin.github.io/ [accessed 2018-04-30] [WebCite Cache ID 6z3yFH15C]
15. Davis L. Instrument review: Getting the most from a panel of experts. Applied Nursing Research 1992;5(4):194-197.
16. DeVon H, Block M, Moyle-Wright P, Ernst D, Hayden S, Lazzara D, et al. A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh 2007;39(2):155-164. [doi: 10.1111/j.1547-5069.2007.00161.x] [Medline: 17535316]
17. van Kuijk J, van Driel L, van Eijk D. Usability in product development practice; an exploratory case study comparing four markets. Appl Ergon 2015 Mar;47:308-323. [doi: 10.1016/j.apergo.2014.10.007] [Medline: 25480002]
18. U.S. Department of Health & Human Services. Usability Evaluation Basic URL: https://www.usability.gov/what-and-why/usability-evaluation.html [WebCite Cache ID 6z3yi5MBH]
19. Ivory M, Marti A. The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. ACM Computing Surveys 2001 Dec;33(4):470-516.
20. Epstein J, Santo R, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol 2015 Apr;68(4):435-441. [doi: 10.1016/j.jclinepi.2014.11.021] [Medline: 25698408]
21. Bolarinwa O. Principles methods of validity reliability testing of questionnaires used in social health science researches. Niger Postgrad Med J 2015;22(4):195-201. [Medline: 26776330]

Abbreviations

CVI: content validity index
FVI: face validity index
SKAMA: Skala Kebolehgunaan Aplikasi Mudah Alih
SUS: System Usability Scale

Edited by G Eysenbach; submitted 06.03.18; peer-reviewed by MSB Yusoff, S Mohd Hairon; comments to author 04.04.18; revised version received 16.04.18; accepted 19.04.18; published 14.05.18.

Please cite as:
Mohamad Marzuki MF, Yaacob NA, Yaacob NM
Translation, Cross-Cultural Adaptation, and Validation of the Malay Version of the System Usability Scale Questionnaire for the Assessment of Mobile Apps
JMIR Hum Factors 2018;5(2):e10308
URL: http://humanfactors.jmir.org/2018/2/e10308/
doi:10.2196/10308
PMID:29759955

Original Paper

Relationship Between Evidence Requirements, User Expectations, and Actual Experiences: Usability Evaluation of the Twazon Arabic Weight Loss App

Aroub Alnasser[1], BSc (Hons), MSc; Janet Kyle[2], BSc (Hons), MSc, PhD; Abdulrahman Alkhalifah[1], MSc, PhD; Debbi Marais[3], MSc, PhD

[1]Food Science and Nutrition Department, College of Food and Agriculture Sciences, King Saud University, Riyadh, Saudi Arabia

[2]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

[3]Warwick Medical School, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Aroub Alnasser, BSc (Hons), MSc
Food Science and Nutrition Department
College of Food and Agriculture Sciences
King Saud University
PO Box 86683
Riyadh, 11632
Saudi Arabia
Phone: 966 118056476
Email: aroub@ksu.edu.sa

Abstract

Background: Saudi Arabia has faced a steady growth in the prevalence of obesity. The concurrent and ubiquitous use of mobile technology, such as smartphones and apps, provides an opportunity for the implementation of mHealth technology, a method for delivering behavioral interventions. Despite their effectiveness in promoting lifestyle and diet modification, culturally adapted weight loss apps and related interventions are lacking in Gulf Cooperation Council countries.

Objective: The objective of our study was to identify the relationship between adherence to evidence-informed practices, potential user expectations, and actual user experiences in order to enhance the understanding of the overall usability of the Twazon Arabic weight loss app.

Methods: In 2 previous studies, 39 Saudi women were recruited for focus group discussions and 240 Saudi women were recruited for an app-based weight loss intervention. Usability of the Twazon Arabic weight loss app was evaluated by analyzing the opinions and experiences of 26 participants who engaged with the Twazon app for 4 months; the System Usability Scale (SUS) and word clouds were used. The results were triangulated with potential user expectations obtained in the focus group discussion and with the findings from an Arabic app screening for evidence-informed practices.

Results: The average reported SUS score was 69.3. The most favored features were the calorie counter, step counter, and physical activity calorie counter. The features in need of improvement were the social network, notifications, and the Twazon Saudi Food Database. Twazon users preferred and found useful 7 of the 13 evidence-informed weight loss practices that were integrated into the features of the app.

Conclusions: Triangulation identified the most notable relationship to be the disparity between user experience and 2 of the evidence-informed practices, namely a minimum weight loss goal of 0.5 to 1 kg/week and social support; no relationship was found between user expectations and evidence-informed weight loss practices. The overall usability of the Twazon Arabic weight loss app ranged between high marginal and acceptable, indicating that some improvements to the app should be considered for implementation in future app-based weight loss interventions of this kind.

(JMIR Hum Factors 2018;5(2):e16) doi:10.2196/humanfactors.9765

KEYWORDS

mHealth; weight loss; obesity; smartphones; mobile applications; Saudi Arabia; women's health

Introduction

It is no longer news that obesity is a problem in Gulf Cooperation Council countries such as Saudi Arabia, affecting more women than men on average. A major driver of this is unhealthy behaviors such as physical inactivity, overeating, and unhealthful food choices [1]. Due to the severity of the epidemic, it is necessary to implement various treatment strategies and conduct interventions that are accessible to a larger population and effective over the long term. As a novel manner in which to deliver behavioral interventions that might be effective in lifestyle and diet modification, implementation of health-related technology, or rather mHealth, has been of emerging interest.

mHealth is a type of electronic health support that is defined as medical and public health practices that are promoted by mobile devices, such as smartphones, patient monitoring devices, personal digital assistants, and other wireless devices [2]. Commercial weight loss apps have been reported to be more engaging than those that are evidence-informed [3]; however, the quality of the information given by the commercial apps is often rated as low [4,5]. It follows that a more comprehensive user-centered design approach [6] that is based on evidence-informed practices, as well as user expectations and experiences, is vital to ensuring the efficacy of mHealth interventions.

Due to the widespread use and accessibility of mobile technology in Saudi Arabia [7], smartphone apps offer a substantial opportunity to support health behavior change and weight management. However, none that are evidence-informed and culturally adapted are available in the region. With the goal of implementing a 4-month weight loss intervention in Saudi Arabia (AA et al, unpublished data, 2017), the Twazon Arabic weight loss app [8] was developed based on the aforementioned factors in addition to behavior strategies [9], such as self-monitoring. To ensure the proper implementation of a complex intervention [10] involving a website or mobile app, *usability* —or how effectively, efficiently, and satisfactorily a user can interact with a user interface [11]—must be investigated.

The Twazon app was designed to be used autonomously by individuals (male or female) who have weight issues, but are otherwise healthy; it is not intended to be used as treatment in a health care system. The prevalence of overweight and obesity among Saudi women, and a scarcity of research done for this demographic, justify the need for a public health intervention to be carried out for women in this region. In this study, we aimed to identify the relationship between adherence to evidence-informed practices, potential user expectations, and actual user experiences in order to enhance the understanding of the overall usability of the Twazon weight loss app. A triangulation analysis revealed the relationship and tensions found between these aspects of the app's components, and the results we report here reflect their compliance with the Twazon app.

Methods

Design Phase: Evidence Requirements and User Expectations

Weight loss apps in general have been found to be lacking in evidence-informed practices, and the majority that are available are commercial and in English. Due to a complete lack of a systematic reviews of weight loss apps in the region, the first step in designing Twazon [8] involved screening 65 Arabic weight loss apps for their adherence to evidence-informed practices [12] as recommended by various health authorities [13-15].

To further inform the development of the Twazon app, a qualitative study was conducted comprising 4 focus group discussions with the goal of determining potential users' preferences and expectations in a weight loss app. A total of 39 Saudi women with overweight and obesity in Riyadh, Saudi Arabia [16] gave oral responses, which were transcribed and translated from Arabic into English by a certified bilingual translator. Discussions were thematically analyzed and categorized for each of the main topics, and specific quotations were identified to correlate with the theme in mind.

Implementation Phase: App Development and Intervention

The result of the app screening and the focus group discussions was the selection of 13 individual evidence-informed practices, which were grouped as follows: weight assessment and goal setting, healthy diet, physical activity, self-monitoring, and social support (see Multimedia Appendix 1). The behavioral strategy of self-monitoring translated to features of the Twazon app that enable users to track their progress.

The Twazon app requires a single log-in, must be connected to the Internet to function properly, and continues to work in the background. It is not a commercial app; it was developed and made freely available to the public through the iTunes (Apple Inc) and Google Play (Google LLC) stores. Daily physical activity, by activity and time spent doing it, is calculated with data from the user-updated physical activity journal and the integrated pedometer; daily water and energy intake are calculated based on user-updated input of consumption (Figure 1 shows the Twazon app interface).

The dashboard provides automatized, individually tailored, user-specific information regarding daily activity, consumption, and goal tracking, which is reset at the beginning of each day by an automatic algorithm. The food palm gives a personalized biweekly graphic display of the user's healthy lifestyle self-assessment score, including physical activity tips when physical activity goals are not met. This feature is also designed to give instant feedback to users if they exceed their daily energy intake goal. The educational tool is used for menu planning, and the food label tips are used to understand the nutritional content of foods consumed. The Twazon app also offers social support, accessible at the bottom of the interface. This social network, which is restricted to users, encourages individuals to share personal health achievements with one another through

the posting and liking of images and text; no other human contact or feedback from the developers is provided.

Postintervention Phase: App Use and Usability

Engagement was based on app use, which was calculated by an automatic algorithm that grouped the participants according to the frequency of user input (AA et al, unpublished data, 2017). This was a necessary step in assessing usability in that only those participants who regularly updated their information could be considered.

Generally, usability testing conducted with 5 participants will identify at least 85% of usability problems [17]; in this study, a sample of 26 users was deemed to be more than sufficient. Participants were asked to assess the overall usability of the app during an individual interview at a predetermined location for 10 to 15 minutes. The overall usability score of the Twazon app was measured using the 10-question System Usability Scale (SUS), which generates a SUS score ranging from 0 to 100 that is associated with a 7-point adjective rating scale: worst imaginable (12.5), awful (20.3), poor (35.7), ok (50.9), good (71.4), excellent (85.5), or best imaginable (90.9) [18]. To determine what is an "acceptable" SUS score for a product to have, or rather whether a product requires more attention and continued improvement, the score is further classified by acceptability ranges—that is, not acceptable (0-50), low/high marginal (51-69), and acceptable (70 and above; Figure 2); higher product acceptability means fewer usability difficulties experienced by a user [19].

Figure 1. Twazon app interface (from top left, clockwise): Twazon dashboard, educational tool; food label tips, and healthy food palm.

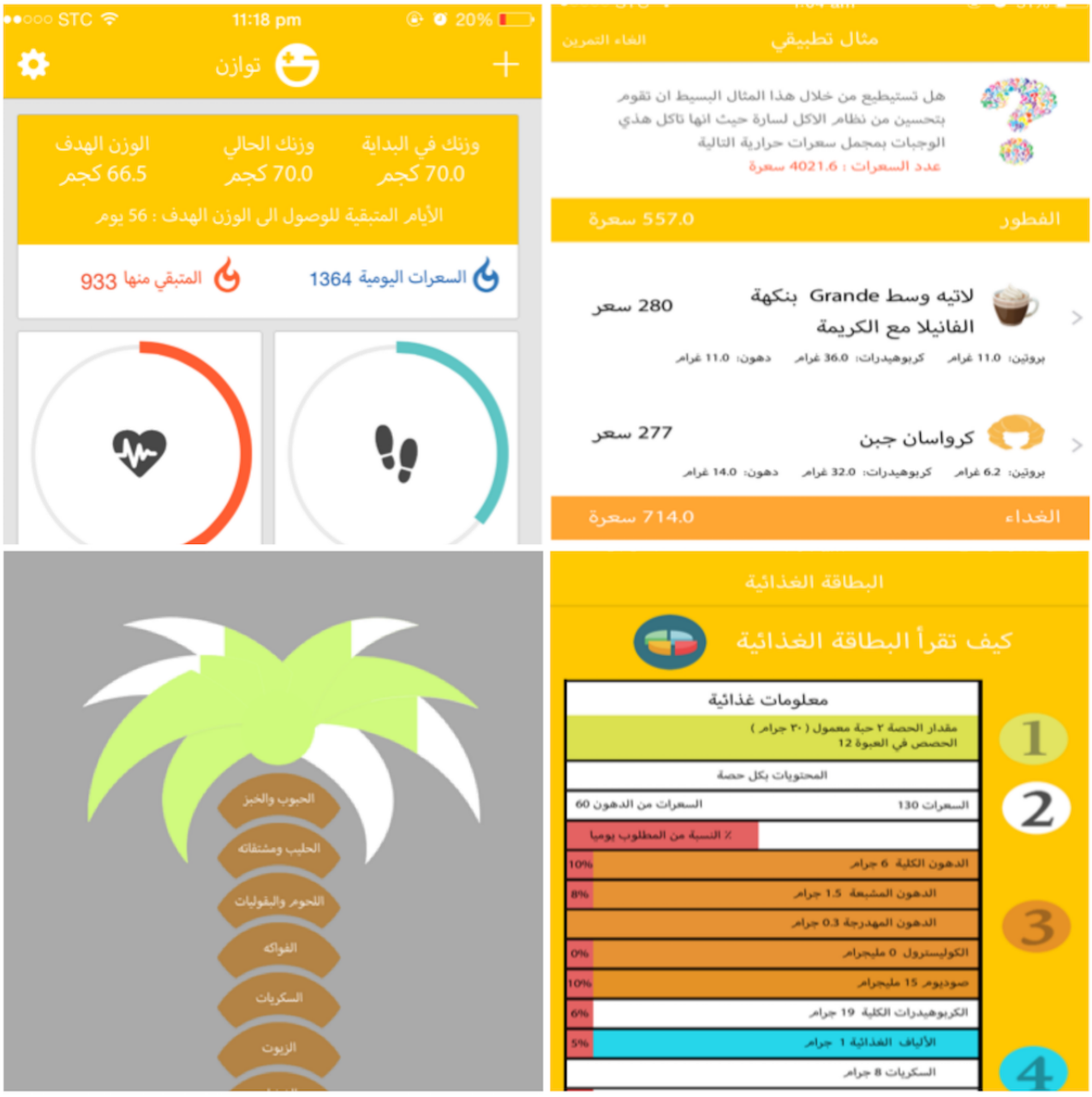
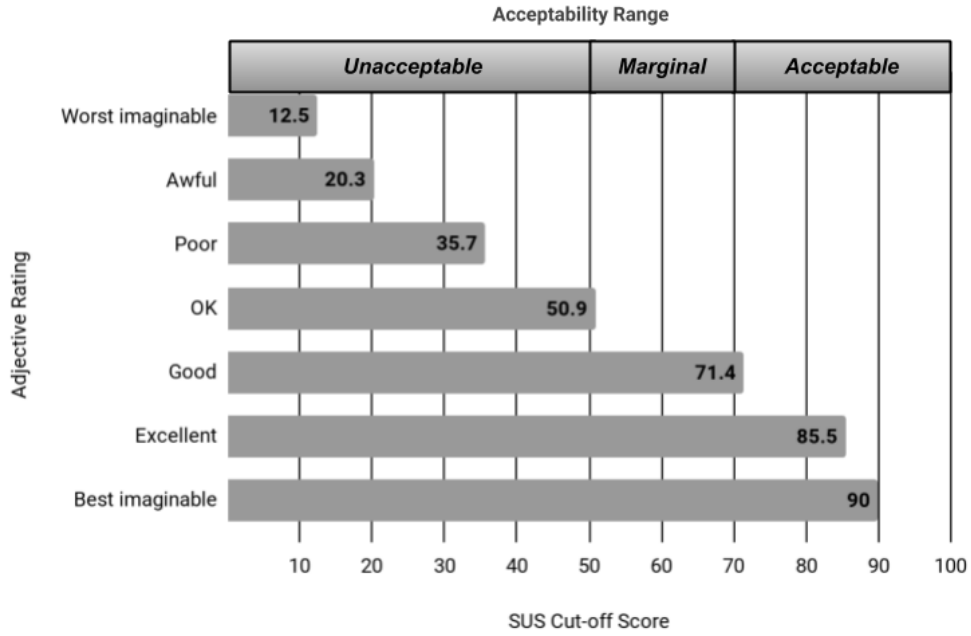


Figure 2. Grade rankings of System Usability Scale (SUS) scores. Adapted from Bangor et al [18,19].



The SUS questions were ranked according to a 5-point Likert scale [20]. Each of the 10 questions had a score range set from 0 to 4. For responses 1, 3, 5, 7, and 9, the score was calculated by subtracting 1 from the scale value. For responses 2, 4, 6, 8, and 10, the score was calculated by subtracting the scale value from 5. The overall score was the total of the scores multiplied by 2.5 [18]. To identify which features of the app participants believed to be the most efficient and which could be improved upon, the participants were asked 2 open-ended questions that were added to the SUS questionnaire but analyzed separately: “What part of the app do you feel works the best?" and “Are there any parts of the app that you feel could be improved, and how?” The results obtained from the 2 additional questions were used to generate visual representations (word clouds) with free online software (Wordle [21]. Word clouds has predominantly been found in social and commercial settings; however, studies have shown that their use in analysis provides “a rapid and practical way to analyse textual data” and helps in “reducing the textual data without bias” [22].

This process allows the reader to quickly identify the most commonly used terms or responses in a given text as it entails illustrating a set of related tags or words in which frequency of word use is reflected visually through font size [23]; this represents the number of participants who gave a response, rather than the total number of responses, and is vital to eliminating the possibility of the same or similar comments being counted multiple times during individual interviews. The answers were sorted into 2 groups (app preferences and app improvements) and a word cloud was generated showing common themes for each question, for a total of 2 word clouds.

The collective results were prepared for analysis in a cross-comparative table using the 13 evidence-informed practice requirements. The information collected in the screening phase was used as a basis for identifying whether these linked to user expectations or experiences.

Results

The Twazon weight loss app intervention was completed by 40 Saudi women with overweight or obesity over the course of 4 months; the rate of attrition was 83%. For the analysis that follows, only the data for the engaged participants (n=26) were used.

System Usability Score (Twazon Intervention)

The overall mean SUS score was 69.3 (SD 10.1), equating to an adjective rating of ok (average=50.9), which suggests that the participants found the app to be more than satisfactory. When compared with the averages for each adjective rating, however, this study’s scores were closer to a rating of good (average=71.4); this translates to an overall acceptability that ranges between high marginal and acceptable (see Figure 2). The highest-rated positive statements responses were numbers 7 (“I imagine most people would learn to use this app very quickly.”) and 9 (“I felt very confident using the app.”). The lowest-rated negative statements were numbers 10 (“I needed to learn a lot of things before I got going with this app.”) and 6 (“I thought there was too much inconsistency in this app.”); see Figure 3.

Word Clouds (Twazon Intervention)

We generated 2 word clouds for the responses given for the 2 open-ended questions regarding the features that were most preferred (question 1) and those that were in need of improvement (question 2). The results for question 1 (Figure 4) showed that the most favored features of the app were the calorie counter, followed by the physical activity calorie calculator and the step counter. The water counter was the fourth most favored feature. The results for question 2 (Figure 5) showed that the primary suggested improvements were to have more food items, followed by change nothing, and then to add more reminders, arrange food items into groups, and social network development.

Figure 3. Mean System Usability Scale (SUS) scores corresponding to the 10 questions. Odd-numbered questions indicate a positive response, while even-numbered questions indicate a negative response. Higher numbers indicate increasing degrees of participant agreement.

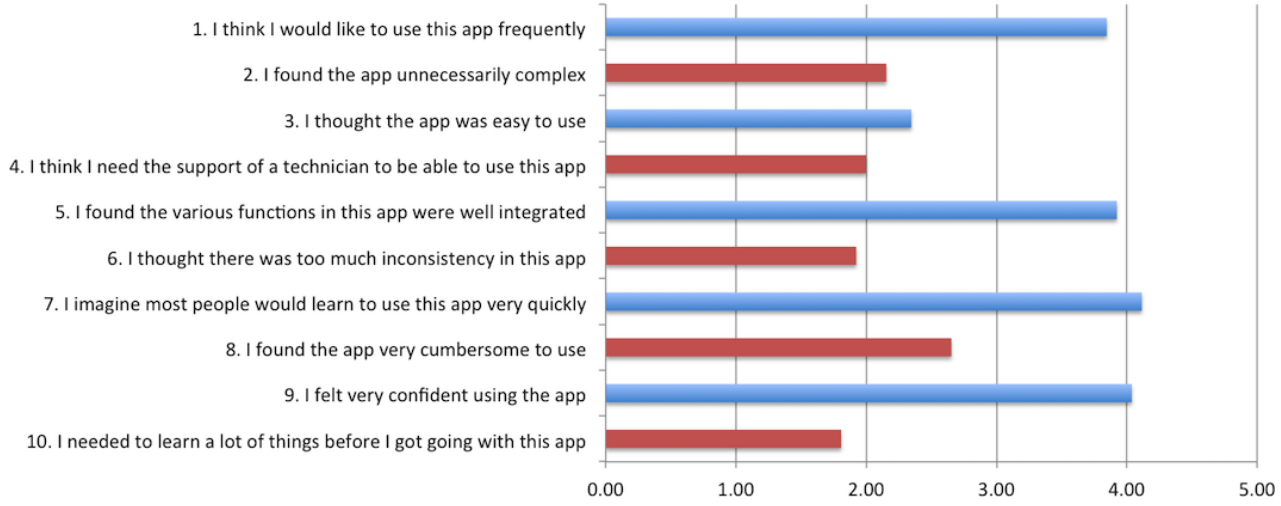


Figure 4. Word cloud representation of responses indicating the most preferred features.

Figure 5. Word cloud representation of responses indicating features in need of improvement.

Triangulation

The results of the screening indicated that Arabic weight loss apps had a very low adherence to evidence-informed practices (median=1); no apps had more than 6 evidence-informed practices, and only 9 apps had 4 to 6 integrated, which justified the need to develop an evidence-informed Arabic app. The focus group discussions then led to exploring potential users' expectations of an ideal app. The results from those discussions indicated that the participants expected all 13 evidence-informed practices to be present in some feature of an ideal app, in addition to its being culturally adapted in terms of language, food, and exercise options (Table 1).

The results obtained from the SUS questionnaires and word clouds indicated that the Twazon users preferred and found useful more than half (7/13) of the evidence-informed practices that were integrated into the apps' features. The participants reported that 2 of the 13 practices were insufficient (weight loss goal of 0.5-1 kg/week and social network), while they did not mention the remaining practices. A cross-comparative analysis (Table 1) highlighted the relationship between adherence to evidence-informed practices and overall usability of the Twazon weight loss app.

Table 1. Cross-comparative analysis of triangulation exploring potential users' expectations of an ideal app. N/A: not applicable. NC: no comment.

Evidence-informed practices	Arabic weight loss apps (adherence; n=65), n (%)	Focus group discussion (expectations)	Twazon intervention (experiences)
1. Meal planning	25 (38)	Yes	Yes
2. Assessing your weight	17 (26)	Yes	NC
3. Regular physical activity	13 (20)	Yes	Yes
4. Maintaining calorie balance	10 (15)	Yes	Yes
5. Keeping a food diary	9 (14)	Yes	Yes
6. Portion control	0 (0)	Yes	NC
7. Eating a diet rich in fruits and vegetables	7 (11)	Yes	Yes
8. Tracking your weight	6 (9)	Yes	NC
9. Keeping a physical activity journal	6 (9)	Yes	Yes
10. Weight loss goal of 0.5-1 kg/week	5 (8)	Yes	No
11. Social support	2 (3)	Yes	No
12. Reading nutrition facts labels	2 (3)	Yes	NC
13. Water instead of soda/juice	7 (11)	Yes	Yes
Additional: culturally sensitive	N/A	Yes	NC
Additional: notifications	N/A	N/A	No

Discussion

Principal Findings

The Twazon app aimed to fill a gap in the research and development of evidence-informed Arabic weight loss apps [12] and interventions in order to find the optimum balance between evidence requirements and user needs. The participants' experiences with the Twazon intervention provided insight into the features of the app that were the least interesting, the most effective, or in need of improvement. When considering the results from the triangulation analysis, a relationship emerged between what is perceived as the best or required practice from the evidence and what participants actually experienced and reported as being useful. Although the Arabic apps failed in general to meet requirements for all 13 evidence-informed practices, the women who took part in the focus group discussions [16] clearly communicated their expectation that all of them should be integrated into an ideal weight loss app.

Some of the evidence-informed practices, such as assessing one's weight and tracking one's weight, were not featured in the word clouds as being favored or in need improvement; this could be attributed to their essential nature in general weight loss programs and apps. The practice of eating a diet rich in fruits and vegetables and the practice of reading nutrition facts labels were also recommended by the app; neither was reported as favored or in need of improvement. Analysis of the SUS score for the question regarding the consumption of fruits and vegetables (AA et al, unpublished data, 2017), however, showed that the participants were successful in increasing their intake. This suggests that the app was effective in promoting this diet modification and practice.

Portion control was also recommended by the app, but the results from the word clouds gave no indication that this was either favored or in need of improvement. This could be attributed to a lack of typical serving sizes, which are found in other countries or in other databases. The development of the Twazon app included the creation of the Twazon Saudi Food Database (AA et al, unpublished data, 2017) with the goal of providing users with a detailed list of household measurements for local and international foods to help promote portion size awareness. However, the portion control feature was not mentioned by participants, implying a need for further investigation into the most effective manner in which it should be implemented.

The most commonly user-reported preferences and proposed improvements suggested that the users were more satisfied with the functions of the app (eg, counters) than with the content (eg, missing food item information).

Primary Preferences

The results showed that the most favored features of the app were related to counters (Figure 4). The 4 evidence-informed practices that fulfill the reported preferences for most favored features are maintaining calorie balance (calorie counter), engaging in regular physical activity (step counter), keeping a physical activity and food journal (physical activity calorie calculator and calorie calculator), and drinking water instead of soda or juice (water counter). Our findings contrast with a recent qualitative study of 24 volunteers that suggested that counters are generally not preferred [24].

These results could be due to the fact that weight loss apps, and more specifically Arabic language apps that are culturally adapted, are relatively new to the region [25] and may be considered a novelty. A quote from one of the participants using the Twazon app illustrates this: "I was using an English weight loss app and against my better judgement I opted to eat pizza and burgers instead of kapsa or jarish so that I could count my calories with this app." The act of counting calories may have

been preferred in this study as a result of the participants' interest in being able to log foods that they were familiar with due to their accessibility in the Twazon Saudi Food Database (AA et al, unpublished data, 2017). In future app development for the Gulf region, counters such as those found in the Twazon app could potentially be useful, as long as the practice remains novel.

Primary Improvements

The primary suggested improvements were to have more food items, followed by have more reminders, arrange food items into groups, and social network development (Figure 5). The second most reported improvement was to change nothing; possible reasons for this were that the users were satisfied with the app, they found the app to be better than their SUS score suggested, or that they simply didn't report accurately. The suggestions for improvements offer an opportunity to reevaluate the features and structure of the app, with the aim being to inform future app development.

Improvements were mentioned in regard to having access to more food items that are ideally arranged into specific groups to allow users to better log their daily consumption. The users' adoption of evidence-informed practices, such as meal planning and portion control, may have been hindered by not being able to enter or find certain foods with ease. However, the participants' inability to report their energy intake could be attributed to a falsely perceived lack of information. In some instances, participants were entering misspelled food items, causing duplicates, or were entering lengthy descriptions of dishes instead of simple keywords; this complicated the task and made logging foods more demanding.

Food data input challenges could be overcome with the addition of a barcode feature, which was one of the three least-reported suggested improvements (see Figure 5). One qualitative study [26] showed that a barcode feature should be considered, as it might improve users' overall opinion of the quantity and types of food items available, enabling users to update their food intake with the ease of scanning food labels. Future apps might then consider expanding the Twazon Saudi Food Database to include more foods, integrate an autocorrect feature for spelling issues, and offer a barcode scanner to simplify food data input. Further investigation into these features and the user's perception of them is needed to test their efficacy prior to carrying out an intervention.

The Twazon app provided users with three different types of notifications: (1) tailored tips based on unmet goals in food groups and physical activity, (2) general tips for foods to consume and foods to avoid, and (3) a reminder to enter weight and fill in the food palm tree assessment [8]; we gave them the option to choose how often (every 2 days, every 3 days, and every week) they received the first and second types of notifications, but the third type was automatically delivered every 2 weeks on completion of the required input. Although this was done to avoid overwhelming the participants, one study by Freyne et al [27] found that 3 notifications daily did not frustrate the users, exemplifying that an increase in notifications is not necessarily a hindrance. We suggest that more communicative contact should be considered in the development of future weight loss apps so as to encourage users to record as much as possible.

The evidence-informed practice of having a minimum weight loss goal of 0.5 to 1 kg/week perhaps identified the greatest relationship; many women reported losing interest in participating in the intervention due to not being satisfied with the aforementioned goal. This outcome could be explained by an aversion to goals perceived as being impossible or unsatisfying [28]. Several studies showed that participants with obesity are not motivated by an overall weight loss goal of 5% to 10%, as is recommended by health professionals, but rather a weight loss goal of 22% to 34% [29-31]. This failure to meet the expectations of patients with obesity suggests that there needs to be a smaller disparity between actual and expected outcomes. If this is achieved, then the probability of negative effects that are seemingly caused by unmet expectations can be lessened, and in turn more positive weight loss outcomes [32] can be achieved.

Evidence-informed weight loss programs have suggested that social networking could have a positive effect on weight loss outcomes; social media-based reports and sharing via social media sites such as Twitter are effective in weight loss interventions [33], as they can help motivate and empower participants to work harder toward their goals. In the Twazon app, we created an original and private social network that was accessible solely to the users of the app. However, the intervention participants reported the need for more social network development. Despite the remarkably high rate of social media use in the region, it seems that Saudi female participants were not as inclined as expected to share and interact in regard to their weight loss experience within the closed group.

Lack of engagement with the social media aspect of the app could be attributed to the participants not having direct support from family and friends, as is typical on most popular social media sites; the Twazon app was accessible exclusively to registrants of the app. However, one study [34] found no significant differences in a 6-month weight loss intervention between 3 different groups, which included a podcast plus Twitter group. Regardless, our results from the word clouds show that participants desired more social media development, suggesting that the use of social media sites as a tool to help promote weight loss and connect weight loss intervention participants should be considered and optimized in future app-based interventions.

Despite the integration of evidence-informed features into the Twazon app, challenges with retention still arose. To improve retention in future app interventions of this kind, modifications to the social networking feature and an increase in the amount of contact with the user is highly recommended. The reported user experiences also suggest that more consideration needs to be given to establishing weight loss goals that are not demotivating in order to facilitate more successful weight loss outcomes.

Conclusion

Participants deemed the Twazon app to be of acceptable usability. The triangulation analysis revealed the greatest

relationship to be the disparity between user experience and 2 of the evidence-informed practices, namely, a minimum weight loss goal of 0.5 to 1 kg/week and social support. In contrast, user expectations coincided with evidence-informed practices and therefore did not provide any relationship. Once the aforementioned improvements are made, it would be feasible for health care providers to recommend the use of Twazon in weight loss programs that involve behavioral modification strategies. Further in-depth exploration through qualitative study is also needed to better understand the relationship observed so as to appeal to the motivating factors that drive participants toward successful outcomes in their weight loss goals when using weight loss apps.

Acknowledgments

This research project was supported by a grant from the Research Center of the Female Scientific and Medical Colleges, Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tools in Twazon addressing evidence-informed weight loss practices.

[PDF File (Adobe PDF File), 34KB - humanfactors_v5i2e16_app1.pdf]

References

1. ALNohair S. Obesity in gulf countries. Int J Health Sci (Qassim) 2014 Jan;8(1):79-83 [FREE Full text] [Medline: 24899882]
2. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
3. 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]
4. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
5. Nikolaou CK, Lean ME. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2017 Dec 13;41(1):200-202 [FREE Full text] [doi: 10.1038/ijo.2016.186] [Medline: 27780974]
6. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
7. KSA ICT indicators end of Q3 2016. Riyadh, Kingdom of Saudi Arabia: Communications and Information Technology Commission; 2016. URL: http://www.citc.gov.sa/en/reportsandstudies/indicators.PDF [accessed 2017-12-28] [WebCite Cache ID 6w2F8dC3Q]
8. Alnasser A, Sathiaseelan A, Al-Khalifa A, Marais D. Development of 'Twazon': an Arabic app for weight loss. JMIR Res Protoc 2016 May 16;5(2):e76 [FREE Full text] [doi: 10.2196/resprot.5497] [Medline: 27185568]
9. Azar KMJ, 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]
10. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Dec 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
11. US Department of Health and Human Services. What & why of usability: glossary. Washington, DC: US Department of Health & Human Services; 2018. URL: https://www.usability.gov/what-and-why/glossary/u/index.html [accessed 2018-03-03] [WebCite Cache ID 6xdqATWbX]
12. Alnasser AA, Amalraj RE, Sathiaseelan A, Al-Khalifa AS, Marais D. Do Arabic weight-loss apps adhere to evidence-informed practices? Transl Behav Med 2016 Dec;6(3):396-402 [FREE Full text] [doi: 10.1007/s13142-015-0347-7] [Medline: 27528528]
13. Centre for Public Health Excellence at NICE. Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children. National Institute for Health and Clinical Excellence 2006 Dec. [Medline: 22497033]
14. Logue J, Thompson L, Romanes F, Wilson DC, Thompson J, Sattar N, Guideline Development Group. Management of obesity: summary of SIGN guideline. BMJ 2010 Feb 24;340:c154. [Medline: 20181637]
15. Australian Government, National Health and Medical Research Council, Department of Health and Ageing. Clinical practice guidelines for the management of overweight and obesity in adults, adolescents and children in Australia. Canberra, Australia:

Commonwealth of Australia; 2013 May. URL: http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/n57_obesity_guidelines_130531.pdf [accessed 2017-12-28] [WebCite Cache ID 6w2FLgch6]
16. Alnasser AA, Alkhalifa AS, Sathiaseelan A, Marais D. What overweight women want from a weight loss app: a qualitative study on arabic women. JMIR Mhealth Uhealth 2015 May 20;3(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4409] [Medline: 25993907]
17. Nielsen J. Why you only need to test with 5 users. Fremont, CA: Nielsen Norman Group; 2000 Mar 19. URL: http://www.useit.com/alertbox/20000319.html [accessed 2017-12-28] [WebCite Cache ID 6w2EnBvlz]
18. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
19. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008;24(6):574-594.
20. Brooke J. SUS-a quick and dirty usability scale. In: Patrick W, Jordan B, Thomas IL, editors. Usability Evaluation In Industry. London, UK: Taylor & Francis; 1996:189-194.
21. Feinberg J. Wordle. 2014. URL: http://www.wordle.net/ [accessed 2018-03-27] [WebCite Cache ID 6yEeLp4ll]
22. Gill D, Griffin A. Good medical practice: what are we trying to say? Textual analysis using tag clouds. Med Educ 2010 Mar;44(3):316-322. [doi: 10.1111/j.1365-2923.2009.03588.x] [Medline: 20444063]
23. Prochaska JJ, Fromont SC, Wa C, Matlow R, Ramo DE, Hall SM. Tobacco use and its treatment among young people in mental health settings: a qualitative analysis. Nicotine Tob Res 2013 Aug;15(8):1427-1435 [FREE Full text] [doi: 10.1093/ntr/nts343] [Medline: 23322765]
24. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
25. Mansar SL, Jariwala S, Shahzad M, Anggraini A, Behih N, AlZeyara A. A usability testing experiment for a localized weight loss mobile application. Procedia Technol 2012;5:839-848. [doi: 10.1016/j.protcy.2012.09.093]
26. 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]
27. Freyne J, Yin J, Brindal E, Hendrie GA, Berkovsky S, Noakes M. Push notifications in diet apps: influencing engagement times and tasks. Int J Hum Comput Interact 2017 Jan 11;12:1-3. [doi: 10.1080/10447318.2017.1278896]
28. Foster GD, Wadden TA, Phelan S, Sarwer DB, Sanderson RS. Obese patients' perceptions of treatment outcomes and the factors that influence them. Arch Intern Med 2001 Sep 24;161(17):2133-2139. [Medline: 11570944]
29. Foster GD, Wadden TA, Vogt RA, Brewer G. What is a reasonable weight loss? Patients' expectations and evaluations of obesity treatment outcomes. J Consult Clin Psychol 1997 Feb;65(1):79-85. [Medline: 9103737]
30. Jeffery RW, Wing RR, Mayer RR. Are smaller weight losses or more achievable weight loss goals better in the long term for obese patients? J Consult Clin Psychol 1998 Aug;66(4):641-645. [Medline: 9735580]
31. O'Neil PM, Smith CF, Foster GD, Anderson DA. The perceived relative worth of reaching and maintaining goal weight. Int J Obes Relat Metab Disord 2000 Aug;24(8):1069-1076. [Medline: 10951549]
32. Leung KK, Silvius JL, Pimlott N, Dalziel W, Drummond N. Why health expectations and hopes are different: the development of a conceptual model. Health Expect 2009 Dec;12(4):347-360 [FREE Full text] [doi: 10.1111/j.1369-7625.2009.00570.x] [Medline: 19689450]
33. Dahl AA, Hales SB, Turner-McGrievy GM. Integrating social media into weight loss interventions. Curr Opin Psychol 2016 Jun;9:11-15. [doi: 10.1016/j.copsyc.2015.09.018]
34. Kozak AT, Buscemi J, Hawkins MAW, Wang ML, Breland JY, Ross KM, et al. Technology-based interventions for weight management: current randomized controlled trial evidence and future directions. J Behav Med 2017 Feb;40(1):99-111. [doi: 10.1007/s10865-016-9805-z] [Medline: 27783259]

Abbreviations

SUS: System Usability Scale

Edited by G Eysenbach; submitted 04.01.18; peer-reviewed by S Barbieri, JY Kim; comments to author 23.01.18; revised version received 03.03.18; accepted 19.03.18; published 17.04.18.

Please cite as:
Alnasser A, Kyle J, Alkhalifah A, Marais D
Relationship Between Evidence Requirements, User Expectations, and Actual Experiences: Usability Evaluation of the Twazon Arabic Weight Loss App
JMIR Hum Factors 2018;5(2):e16
URL: http://humanfactors.jmir.org/2018/2/e16/
doi:10.2196/humanfactors.9765
PMID:29666042

Original Paper

A Novel Information Retrieval Tool to Find Hospital Care Team Members: Development and Usability Study

Kyle Morawski[1*], MPH, MD; Craig Monsen[1*], MD; Sukhjit Takhar[2*], MPH, MD; Adam Landman[2*], MS, MIS, MHS, MD

[1]Atrius Health, Boston, MA, United States

[2]Emergency Medicine, Brigham & Women's Hospital, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Kyle Morawski, MPH, MD
Atrius Health
133 Brookline Ave
Boston, MA,
United States
Phone: 1 617 421 1000
Email: kyle_morawski@atriushealth.org

Abstract

Background: Hospital communication among members of a patient's care team is a central part of clinical workflow and consumes a large amount of a health care provider's time. Oftentimes the complexity of hospital care leads to difficulty in finding the appropriate contact, which can lead to inefficiencies and frustration. Squire is a Web-based information retrieval app created to improve the speed and efficiency in reaching the appropriate team member during the care of a hospitalized patient.

Objective: The objective of the study was to design and develop Squire and to evaluate the usage, usability, and perceived effect of the app on finding the correct contact within a hospital.

Methods: We used a mixed-methods design using a before-after survey methodology combined with one-on-one interviews to understand the perceived effect of Squire. The study took place at an academic medical center with internal medicine resident physicians. We surveyed residents on demographics, as well as time and efficiency of hospital communication before and after the use of Squire. After using Squire, participants were also asked to evaluate Squire's Net Promoter Score (NPS). A subset of voluntary participants participated in one-on-one interviews and completed the System Usability Scale (SUS). We performed descriptive statistics on participant characteristics, app usage data, and responses to surveys. Survey results were compared before and after Squire adoption using the Wilcoxon rank-sum test and a general linear model. Interview data were analyzed using content analysis with a qualitative description approach to review and categorize feedback from participants.

Results: There was a 67.9% (74/109) response rate to the pre-Squire survey and 89.9% (98/109) response rate to the post-Squire survey. At baseline, there was an average of 22.2 (95% CI 18.4-26.0) minutes/day spent searching for the right contact, and this decreased to 16.3 (95% CI 13.9-18.7) minutes/day after Squire was launched (P=.01). There were favorable usability scores, with an average SUS of 84.7, and a marginal NPS of +6.1. Overall, the use of Squire included 22,283 page views, most commonly to contact the admissions office or portable chest x-ray technician. Interviews highlighted common benefits of Squire, including decreased perceived time spent on hold with operators and improvement in connecting with the appropriate contact in specialized, complex departments. Future opportunities were also identified to improve Squire including adding a two-way communication between physician and nursing staff and providing offline access.

Conclusions: Squire decreased the perceived time required to find an appropriate contact and had a favorable usability score; however, the NPS was marginal and several opportunities were identified to improve Squire for future use.

(JMIR Hum Factors 2018;5(2):e14) doi:10.2196/humanfactors.6781

KEYWORDS

communication; patient care team

Introduction

Background

The complexity of current medical care requires frequent communication within the care team, but many systems do not allow this to be done efficiently. Most academic medical centers have grown piece-by-piece, rather than being designed to function as a coherent whole. In fact, communication has become so centralized that some hospitals are devoting entire departments to this endeavor [1]. Previous studies have shown that the amount of time spent talking to providers is almost double than that of direct patient care [2]. Patient safety has also been shown to be dependent on good team communication [3], and the economic burden of communication inefficiency has been estimated at US $12 billion per year in the United States [4].

At our hospital, there are 2 main workflows for contacting the most appropriate care team member: (1) one can call the hospital operator and wait to be connected or (2) utilize the hospital's Web-based paging directory and search for the correct contact. Many people find wait times with the operator long and the paging directory difficult to search. Both can be ineffective because of poor matching and unclear role description. While the immediate care team members (attending physician, resident physician, nurse) are listed in the electronic health record (EHR), other care team members such as the respiratory therapist, echocardiogram technician, or radiologist can be more difficult to locate.

Importance

Up to one-fifth of a medical intern's time is used for talking with other providers, representing the single largest activity performed during a workday [2]. In a 2013 study from Johns Hopkins, talking with other providers was more time-consuming than direct patient care, which represented 12.3% of a medical intern's time. As this is a large portion of one's time and care becomes more complex, it will be necessary to optimize how we identify and contact members of a patient's health care team.

Technology has been lauded as a solution to help improve health care delivery efficiency. If the benefits of technology are to be realized, such that the health care system is able to achieve improved value in the setting of expanding complexity of patients and care, there must be a focus on the human factor in care redesign and process flow [3,5]. There can often be unintended consequences of introducing new technologies, therefore evaluating users' response to a new tool is important to ensure that desired positive impacts are achieved [6].

Goals of This Intervention

We designed and implemented a novel Web-based information retrieval app, Squire, to improve the speed and efficiency to reach the appropriate team member during the care of a hospitalized patient at a large academic medical center. With increasing complexity of care, work hour restrictions, and demands for productivity in current hospital medicine, Squire aims to facilitate contacting the correct member of a patient's care team and to reduce the need to call the hospital operator. All interface construction, back-end programming, and user experience was focused on speed of activity completion. In this paper, we describe the design and development of Squire and then evaluate the usage and usability of the platform, as well as its perceived effect on efficiency in finding the correct contact in a real-world setting.

Methods

Study Design and Setting

In this mixed-methods study, we evaluated the Squire app using a before-after survey methodology combined with purposefully selected, semistructured individual interviews. We performed the study in an academic medical center, with internal medicine resident physicians using the app during their usual clinical practice.

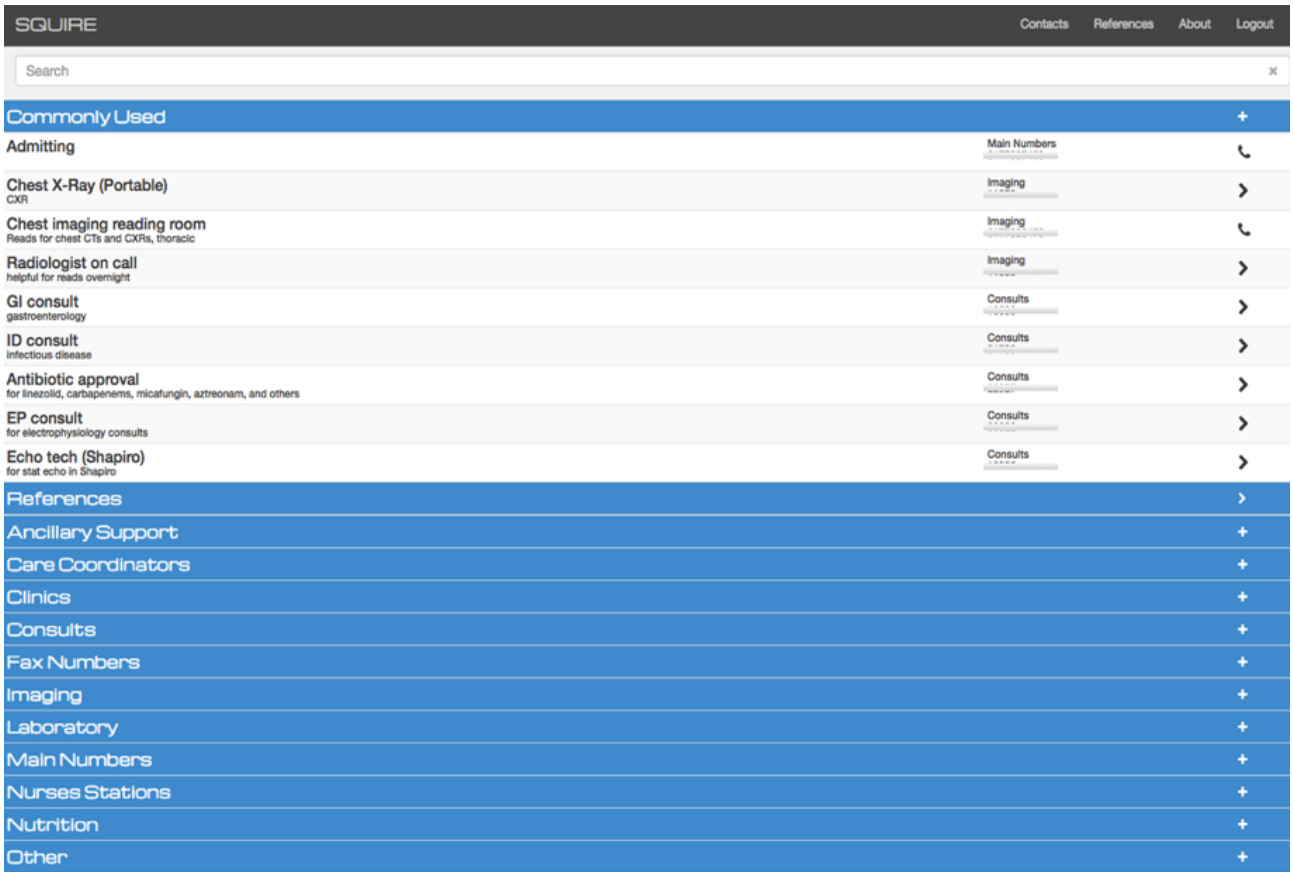
The Partners Health Care institutional review board (Partners Health Care, Boston, MA) deemed this study exempt from review.

Intervention: Squire

Squire is a Web- and mobile-based software app designed to offer clinicians with quick access to commonly used resources, including hospital back office phone numbers, the hospital paging system, and clinical references (Figure 1). Squire was conceived and developed by one of the authors (CM), a dually trained internal medicine physician and clinical informaticist, using an iterative, user-centered design approach [7]. Given CM's expertise with the app context, requirements, and capabilities, he created the initial concept and design. A small cohort of pilot users provided critical feedback including the most useful contact numbers, broken links, and appropriate groupings of contact information. In these early feedback sessions, as well as in previous research [8,9], it was clear that speed and simplicity were of paramount importance. These users also provided iterative feedback on mock-ups and prototypes, leading to interface improvements to optimize usability and satisfy real-world settings [7].

Users log in via computer workstations or mobile phones through the internet, using their hospital clinical system credentials. Phone and pager numbers are listed in a searchable directory. As distinct from existing tools, the directory includes indexed, searchable comments that may be modified based on user feedback in addition to titles, phone numbers, and pager numbers to aid in identifying the most appropriate contact. These contacts include consult services, radiology reading rooms, laboratory departments, nurses stations, pharmacists, care coordinators, and nearby hospitals among others. Users may initiate a call directly from their phone by selecting the contact or sending a text page by selecting a pager contact and entering a message into a structured paging Web form.

Figure 1. The Squire application landing page with most commonly used contacts displayed.



The app is delivered as a website hosted on an internal Partners Health Care CentOS Linux server. The site is accessible on all platforms including mobile using responsive design JavaScript, CSS (cascading style sheets), and HTML5. Responsive design means that the appearance of the website adjusts dynamically to where and how the user is viewing it, responding to features such as the device, browser, and window size. This technique is now ubiquitous for highly trafficked websites. Bootstrap and JQuery, open source frameworks, were used for the front-end user interface using a moderate amount of custom JavaScript and CSS to optimize the experience.

The back end architecture also consists of open source technologies, including Ruby on Rails served using an Apache HTTP server. Data are stored on a SQLite database. It should be noted that there is no personally identifiable or protected health information stored on the server or in the app. Figure 2 summarizes the system's technical architecture.

There are two noteworthy integration points for the software: (1) user authentication and (2) paging. First, we integrate with the hospital's lightweight directory access protocol (LDAP) servers so users can use their hospital clinical systems credentials to log in to Squire. User authentication is performed by the server after a user has entered in their credentials via a secure socket-layer (SSL)-enabled, encrypted LDAP adaptor. This securely checks against the hospital's LDAP servers so that a user with the provided username and password is authorized before allowing access to the app. Second, text pages can be sent directly from the Squire app as a result of integration with our hospital's Paging Directory Service. This is a simple-object access protocol (SOAP)-enabled service through which "3rd party" apps can be built to send pages on the hospital's paging network [10]. It also allows apps to search the directory to match users to pager numbers and to determine which users are currently accepting pages.

Participant Selection

Internal medicine resident physicians at a large academic medical center were provided access to Squire between January 2015 and February 2016. We selected internal medicine residents because they frequently need to identify and communicate with other patient care team members, such as specialty physicians, care coordinators, and respiratory therapists. At this institution, there are 109 residents in internal medicine annually, of which 44% are female, with an average age of 30 (range 25-42) years.

Study Protocol

Before availability of Squire, we emailed the residency with a baseline survey (Figure 3) of their contact searching challenges, including how often they are frustrated by not finding the right person to contact and how much time they spend searching for right contacts each day.

Figure 2. The Squire technical architecture. LDAP: lightweight director access protocol; PPD: partners phone directory; SOAP: simple object access protocol; SSL: security service provider.

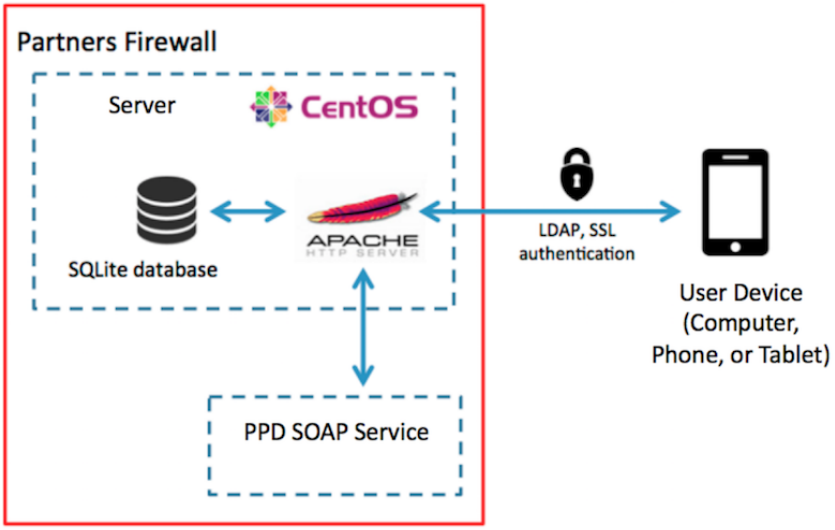


Figure 3. Evaluation survey pre- and post-Squire use. The letter a signifies only present on the post-implementation survey.

1) What best describes you?										
PGY-1	PGY-2	PGY-3	PGY-4 or more							
2) What is your gender?										
male	female									
3) What is your comfort with technology?										
tech-challenged	average comfort with technology	tech-savvy								
4) How often are you frustrated by not finding the right person to contact?										
daily	weekly	monthly	never							
5) How much time do you spend searching for the right contact each day?										
0-4 minutes	5-14 minutes	15-29 minutes	30-59 minutes	>60 minutes						
6) How many days each week do you typically use Squire?[a]										
never	1-3 days/week	3-5 days/week	≥ 5 days/week							
7) How likely is it that you would recommend using Squire to a friend or colleague?[a]										
(unlikely) 0	1	2	3	4	5	6	7	8	9	10 (likely)

Squire was made available to all internal medicine resident physicians in February 2015 by an announcement at a resident conference and sending email notifications. We allowed resident physicians to use Squire for 6 months and then sent an evaluative survey using Research Electronic Data Capture (REDCap, Vanderbilt University, Nashville, TN), a secure, Web-based app designed to support data capture for research studies [11]. The post-Squire survey mirrored the baseline survey, with the addition of Net Promoter Score (NPS) and a question regarding use of Squire in everyday practice (Figure 3, question 7). NPS [12,13] is used by many to evaluate how likely someone is to recommend the new product or technology to a friend or family member. Survey respondents were also asked whether they would be willing to participate in a follow-up one-on-one, semistructured interview about their use of Squire. Participants had the option to stop the survey at any time.

There were 98 responses (90% response rate) to the survey, with 81 of the respondents (83%) indicating acceptance to be interviewed. Survey participants willing to be interviewed were arranged in tertiles with respect to number of log-ins to Squire, and 9 interviewees (10% of total respondents), were purposefully selected [14] from this list, blocking on number of log-ins. The interviewer used a one-on-one, semistructured approach [15] with an interview guide, but with allowance for the interviewee to bring up themes at their discretion (see Figure 4). The interview ended with the System Usability Scale (SUS; see Figure 4, question 6), a validated measure of system usability [16,17]. The interviewer took notes and audio-recorded the interviews. Participation in the one-on-one interviews was voluntary; participation and feedback provided did not impact professional standing or performance evaluations. All qualitative interview participants provided verbal informed consent and were compensated with a US $30 gift card for attending the interview process.

In addition to user experience evaluation described above, we tracked Squire usage statistics through audit logs, including number of log-ins, commonly used features, and total number of users over time.

Figure 4. Semistructured interview guide.

1) How often did you use Squire? (for the group that was in the lowest tertile, a follow up questions of, "Why not?" will be asked)
2) What was the impact of Squire on your clinical practice and efficiency?
3) What problem to you feel Squire has yet to address?
4) Were there any features of Squire that you found frustrating?
5) What other ways would you improve communication and access to information in the hospital?
6) System Usability Scale Participants will be asked to score the following 10 items with one of five responses that range from Strongly Agree to Strongly Disagree: - I think that I would like to use this system frequently. - I found the system unnecessarily complex. - I thought the system was easy to use. - I think that I would need the support of a technical person to be able to use this system. - I found the various functions in this system were well integrated. - I thought there was too much inconsistency in this system. - I would imagine that most people would learn to use this system very quickly. - I found the system very cumbersome to use. - I felt very confident using the system. - I needed to learn a lot of things before I could get going with this system.

Outcome Measures

Overall, we evaluated Squire's usage and usability, as well as the perceived effect of Squire on efficiency of finding the correct contact in the hospital setting. We measured Squire usage through logs of unique users for Squire, frequency of page views overall, and frequency of specific page views to identify most commonly used features. We also surveyed the users on how often they used Squire (Figure 3). Usability was measured by the SUS, NPS, and exploratory, qualitative semistructured interviews with the users. Time spent searching for the appropriate contact was measured before and after Squire implementation by a 5-category Likert scale survey question (Figure 3) with the following intervals: 0-4 min, 5-14 min, 15-29 min, 30-59 min, and >60 min.

Analysis Approach

We present participant characteristics, overall use, and most commonly used features of the Squire app with descriptive statistics. To analyze the survey results before and after the use of Squire, we used the Wilcoxon rank-sum test. Since ordinal category differences can be difficult to interpret, we also performed an adjunct analysis to estimate the average time saved with the Squire platform, an approach supported by prior research [18]. We compared the mean time spent searching for the right contact each day before and after Squire implementation using a general linear model with a link function and robust variance to show magnitude of findings, using each ordinal unit's midpoint [19,20]. For those who spent over 60 min, we used 70 min as the mean time spent searching for the right contact, providing a conservative estimate, minimizing the effect of outliers.

A content analysis [18,21] was performed on the one-on-one interview data using a qualitative description approach. Two of the investigators (KM and CM) reviewed notes and audio recordings, coding and sorting content to identify key phrases and meaningful text units. Both investigators performed this task independently, then met to discuss categories and subcategories of feedback, iteratively revising until consensus was reached. These investigators selected representative quotes for each of the categories identified, extracting quotes from the audio recordings to ensure accuracy.

Qualitative data were managed using Microsoft Excel (Microsoft Corporation, Redmond, WA); quantitative data were analyzed in STATA 14 (StataCorp, LLC, College Station, TX).

Results

Characteristics of Study Subjects

There was a 67.9% response rate (74/109) in the baseline survey, and an 89.9% response rate (98/109) in the follow-up survey. Characteristics were similar between the 2 groups with respect to postgraduate year, sex, and level of comfort with technology (Table 1).

Survey Results

Survey Response

In the baseline survey, 97% (72/74) of respondents felt that they were frustrated by the difficulty in finding the right person to contact either daily or weekly (Table 2). None responded that they were never frustrated by inability to contact the right person. Nearly three-fourth of the respondents felt that they spent 30 min or less a day searching for the right contact, whereas the remainder felt that they spent more than 30 min daily.

After implementation of Squire, we observed a significant decrease (P=.02) in the amount of time spent in finding the right person to contact (Figure 5 and Table 2). In our regression model, we also found that participants spent 5.8 min (95% CI 1.6-10.2) less searching for the right contact each day after Squire implementation. There were still no participants who were never frustrated by trying to find the right person to contact.

Table 1. Characteristics of survey participants pre- and post-Squire. PGY: postgraduate year.

Characteristics of survey participants	Pre-Squire (n=74), n (%)	Post-Squire (n=98), n (%)	*P* value[a]
Resident training level			.28
PGY-1	35 (47)	36 (37)	
PGY-2	20 (27)	37 (38)	
PGY-3	16 (22)	22 (22)	
PGY-4 or more	2 (3)	3 (3)	
Female	42 (57)	38 (39)	.55
Technology comfort level			.15
Tech-challenged	1 (1)	3 (3)	
Average comfort	53 (72)	62 (63)	
Tech-savvy	14 (19)	33 (34)	

[a]*P* value for group differences calculated with Wilcoxon rank-sum test.

Table 2. Comparison of care team communication efficiency pre- and post-Squire and reported use of Squire.

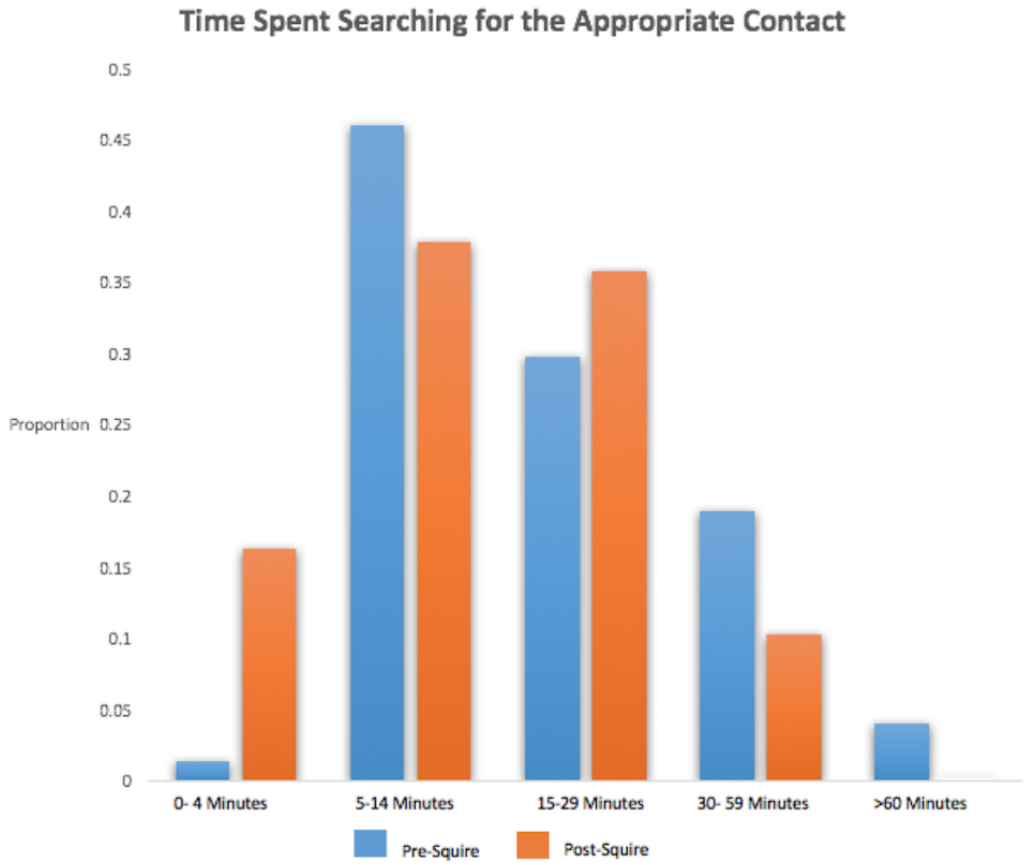
Care team communication	Pre-Squire, n (%)	Total minutes searching per day[a]	Post-Squire, n (%)	Total minutes searching per day[a]	*P* value
How often were you frustrated by not finding the right person to contact?					.66[b]
Daily	40 (54)		58 (59)		
Weekly	32 (43)		34 (35)		
Monthly	2 (3)		6 (6)		
Never	0 (0)		0 (0)		
How much time do you spend searching for the right contact each day?					.02[b]
0-4 mins	1 (1)	2	16 (16)	32	
5-14 mins	34 (46)	323	37 (38)	351	
15-29 mins	22 (30)	484	35 (36)	770	
30-59 mins	14 (19)	623	10 (10)	590	
>60 mins	3 (4)	140	0 (0)	0	
Average time searching for contact per person per day (95% CI)		22.2 (18.4-26.0)		16.3 (13.9-18.7)	.01[c]
How many days each week do you use SQUIRE?					
Never			34 (35)		
1-3 days/week			46 (47)		
3-5 days/week			8 (8)		
>5 days/week			10 (10)		

[a]Midpoint from range of time multiplied by n (ie, midpoint of 5-14 mins is 9.5 mins, multiplied by 34 participants who selected that range, results in a total of 323 minutes searching per day).

[b]*P* value for group differences, calculated with Wilcoxon rank-sum test.

[c]*P* value calculated using general linear model with robust variances.

Figure 5. Time spent searching for appropriate contact pre- and post-Squire use.



A majority (74%, 72/98) reported spending between 5 and 30 min a day searching for the right contact; however, 16% (16/98) reported spending less than 5 min a day searching, equating to an absolute increase of 15% with respect to pre-Squire survey.

Use of Squire was reported as being typically less than 3 times each week by 82% (80/98) of respondents; however, there was a small proportion (10 respondents, 10%) who used it 5 or more times each week.

Net Promoter Score, System Usability Scale

Of the 98 respondents to the postimplementation survey, 32% (32/98) scored the likelihood of recommending Squire to a friend or colleague as 6 or below on a 10 point Likert scale, and thus were classified as detractors. Thirty-nine percent (38/98) of the respondents scored this same question as a 9 or higher and were classified as promoters, and 20% (20/98) respondents provided a score of 7 or 8 and were classified as neutral. This provided an overall NPS of 6.1. The SUS resulted in a mean score of 84.7 on a scale of 0 to 100. The scores ranged from 70 to 97.5.

Most Commonly Used Features

During the 6-month period between launching Squire and performing the evaluation, there were 312 unique users and 22,283 page views. The most commonly viewed features were to contact the admission office (279 views, 2.3% of total views), portable chest x-ray technician (240 views, 1.1% of total views), or the chest imaging reading room (234 views, 1.1% of total views).

Qualitative Interview Results

Participants identified 3 major categories of feedback on Squire during the one-on-one interviews: (1) reducing hold time with hospital operators; (2) value in complex, specialized departments; and (3) opportunities for improvement. Table 3 summarizes these categories and provides additional illustrative participant quotes.

Use of Squire Reduces Time Spent on Hold With Hospital Operators

Seven respondents commented that the largest impact on efficiency in the hospital is with not having to wait on hold with the operator while being transferred to the desired contact. It was cited that this could save, “5 minutes with each call,” and, “has allowed…patients to get more timely care” (Participant 9, Post Graduate Year (PGY) 2). Furthermore, 3 respondents indicated that finding the appropriate number to call was only in Squire and not present with the current Web-based paging directory. Participant 1 (PGY 2) explained, “There are so many headaches during residency, and trying to find the right number shouldn’t be one of them” (see Table 3).

Table 3. Key categories and themes identified during one-on-one interviews with illustrative quotes from participants. PGY: postgraduate year.

Theme	Example quote(s)
Use of Squire reduces time spent on hold with hospital operators	"Could quantify the time it could peel off the day or week." [Participant 1, PGY-2]
	"It just makes everything quicker, I used to wait on hold with the operator, now I can just look it up." [Participant 8, PGY-1]
Squire is particularly valuable for finding contacts from specialized, complex departments	"There are so many headaches during residency, and trying to find the right number shouldn't be one of them." [Participant 1, PGY-2]
	"Could save me 5 minutes depending on how many wrong phone calls I make or get connected to the wrong places." [Participant 9, PGY-2]
Opportunities for improvement of Squire	
Two-way communication with the nursing staff	"There is a lot of 'cat and mouse' with trying to call back [nursing staff], especially on night float." [Participant 3, PGY-1]
	"If in Squire we knew the nurse's name and contact information it would speed things up." [Participant 9, PGY-2]
Offline access	"If you could just take out your smartphone and use the features without waiting for a connection to login that would be great." [Participant 4, PGY-3]

Value for Finding Contacts From Specialized, Complex Departments

Nearly all of participants who were interviewed indicated Squire helped most with finding the appropriate person to call from specialized and complex departments. Two areas that were referenced multiple times were the radiology department and the care-coordination department. In these circumstances, the extra numbers provided in Squire were felt to increase efficiency by requiring less inappropriate calls and redirection to the correct contact. One PGY 2 participant commented that, "It's almost as if [specialty service] wants paging to be frustrating."

Opportunities for Improvement: Two-Way Communication and Offline Mode

There were several areas reported as needing further work. Four interview participants indicated that two-way communication with nursing would be necessary to improve communication efficiency and decrease hold times. They all described instances of being paged by a nurse to the central nursing station and having to wait while the nurse who paged them was found. Many participants also indicated that the need to log in was a barrier to use of Squire. Recommendations for enabling the app function offline (without live network connection) with incremental updates as needed were suggested to improve the usability and efficiency.

Discussion

Principal Findings

We described the design and development of a Web-based information retrieval app, Squire, to improve the speed and efficiency of finding the appropriate contact of a hospital care team. In a pilot with internal medicine resident physicians, 301 users accessed 22,000 page views; however, the majority of users reported only limited use each week. Users reported a strong SUS of 84.6 but a marginal NPS of 6.1. In qualitative interviews, participants provided constructive feedback on features that could be improved. We found that users spent 5.8 min less self-reported time searching for contacts per day after Squire implementation, although there was no change in user frustration levels. While a savings of 5.8 min per day may seem small, when averaged over longer time periods and a population of clinical users, the time savings is substantial.

The most commonly used features were to call the admissions office and the radiology technicians. In general, these are commonly accessed hospital departments but may represent a gap in our institution's current paging directory that does not easily provide these frequently used numbers. These two department numbers are also visible on the front page of the Squire app without any additional searching or scrolling. The qualitative interviews found that Squire was particularly valuable for specialized, complex departments. Radiology is an example of a complex department, with multiple imaging modalities (technicians) and specialty radiology reading rooms (radiologists) spread across a large campus. This inherent complexity and large number of radiology phone number options may also explain why radiology was a commonly used Squire feature.

We expected that Squire would be a frequently used information retrieval tool; however, we found that approximately half of the survey respondents reported that they used Squire 1 to 3 days/week, a larger than expected percentage of respondents (35%) reported that they never used Squire, and only 18% of respondents reported using Squire for 3 or more days/week. These seemingly low reported usage patterns suggest a limit to the value of Squire in everyday clinical practice. Some users may be reserving Squire use for cases in which the phone numbers are difficult to locate via other methods. We also noted that resident physicians rotate roles and call schedules and therefore may have variable need for Squire in any given week. Since we did not specifically ask users to explain their usage frequency, further research is needed to understand Squire usage patterns and whether this reflects limitations of Squire functionality and usefulness.

Squire received a favorable SUS of 84.6, well above the generally accepted average SUS score of 68 [17,22], indicating that Squire was intuitive and easy to learn. Previous research indicates that a score above 82 corresponds to someone being a "promoter" of a new technology [17]. In contrast, Squire's NPS of 6.1 was marginal. Overall, an NPS greater than zero is "good," as positive scores mean that there are more promoters of the product than detractors. Thresholds of 50 have been described as "excellent" and above 70 as "world class." [23]. The Temkin Group benchmarks NPS by industry sectors and found that software had a mean NPS of 41 with a range of 28 to 55 [24]. If we benchmark against health care software, 4 Acute EHRs had NPS of −65, −64, −38, and 0. On balance, Squire's NPS of 6.1 is outstanding compared with EHRs but mediocre when compared with other software companies suggesting an opportunity to improve Squire and guide further iterations over time with serial internal NPS measurements [12].

While Squire's NPS was marginal, we also observed actual promotions of the product to additional users. When Squire was deployed there was no incentive to use it or recommend it to others, yet after 6 months, there were 312 unique users; however, Squire was only rolled out to the 109 internal medicine residents as a part of this study. The observation that app use has naturally diffused outside of the initial study group suggests that there is value to the app outside of the studied individuals and provides some support that positive findings would generalize to clinicians outside of the study population.

A priori, we expected time to search for contacts and user frustration to be correlated. We found that time to search was reduced after the introduction of Squire but frustration levels were not significantly different statistically. Given our small sample size, it is possible that we did not detect a small change in frustration. Furthermore, it is possible that larger time savings are needed to change frustration levels and that we may not have reached these levels with Squire. Qualitative interview feedback confirmed that Squire helped reduce the time that physicians spent waiting on hold for the operator or calling the incorrect contact, but there may also be other factors impacting frustration, such as hold time and redirection to another contact even when the correct phone number is called. Further work is needed here, as efforts to improve the efficiency of nonpatient care activities, have the potential to increase focus for physicians on more critical patient care activities, reduce frustration, and improve the overall efficiency of health care delivery.

Comparison With Prior Work

Entrepreneurial endeavors exist to create a mobile phone app to simplify phone directories [25] or improve access to clinical references [26]; however, research on usability or impact on clinical practice is lacking. There exists previous research regarding development and usability of physician directories [27]; however, effectiveness of implementation remains a poorly studied topic.

Squire was developed to improve efficiency in finding the correct contact among the care team of a hospitalized patient. Mobile usage in the hospital has been increasing, with the main reason cited being speed [28]. In order to integrate mobile devices and new technology into incumbent processes, they must be seamless and represent minimal practice change [13]. Squire attempted to address these issues through working with an already present paging and directory system, allowing for clear descriptions of appropriate numbers to call, and integrating into a mobile interface so that paging and calling can occur directly from a mobile device. Further development is still required in this arena, with this study suggesting that two-way communication would be welcomed by users. This sentiment has been described previously [6] and been shown to improve closed-loop communication [29].

Limitations

There are several limitations to this study that are inherent to its design. First, the study was performed at a single institution with a single group of internal medicine residents, so the generalizability to other institutions and specialties may be limited; however, we had excellent response rates among those requested to participate, which adds to the validity of responses and representativeness of the results to our institution's internal medicine residents. While there were not significant differences between the initial and post-Squire survey participants, there was a slight increase in PGY-2 representation (27% initial and 38% post-Squire). This increase in more experienced survey respondents could contribute to improvements in time to find the correct contact.

We did not capture participant identifiers for the baseline or post-Squire surveys, and therefore it is unknown up to what extent baseline survey respondents are also represented in the post-Squire survey. Furthermore, we were not able to account for the repeated measures in the statistical analysis.

We used a small sample size for the qualitative interviews. Our content analysis approach identified key, common themes; however, it is possible that additional concepts or themes would have emerged with additional participants.

The launch of Squire coincided with the implementation of a new EHR system in our hospital, which may have impacted residents' self-assessment of efficiency. Previous research supports that physicians are more likely to lose optimism, increase time entering orders, and increase overall work time after implementation of an EHR [30]. As this was a study of perception about inefficient time, the concomitant EHR change could have undermined efficiency improvements from Squire. It is also possible that the EHR may have improved efficiency, confounding the results in the opposite direction; however, results from the interviews suggest that users attributed the noted efficiency gains to use of Squire.

Squire was custom developed and is currently available only in our institution. We anticipate that other institutions have similar challenges finding the most appropriate contact and therefore included substantial technical implementation details in the Methods section so other institutions could replicate Squire if desired.

Finally, we used a self-reported outcome of time spent searching for appropriate contacts that could be biased; a more direct measurement of this time may be more accurate. A future time-motion study could provide more robust measures of Squire's impact on efficiency. Furthermore, in our regression

model we collapsed the self-assessed outcome of time with unequal intervals into a mean time. Since these are in any case self-reported times, we do not believe this changes the results appreciably.

Conclusions

We developed a Web-based information retrieval app, Squire, and found that its use saved a modest amount of time per day searching for the correct contact in a hospital setting. While users also found the system highly usable, Squire did not improve the frustration in finding appropriate contacts, and the NPS was a mediocre 6.1. We also identified opportunities to iteratively improve Squire's usability and features. While the study results were mixed, Squire has shown some value in improving the efficiency of finding the appropriate hospital care team member. As we iterate Squire based on the study findings, we have started extending Squire to other user groups and use cases. Squire may also be of interest to other institutions, so we described the technical design so that others can replicate Squire.

Acknowledgments

This study was performed with funding from a Martin P Solomon grant provided by Brigham and Women's Hospital. The authors are responsible for the quality of the data and analysis. We would like to acknowledge the internal medicine residents who have used Squire and provided feedback about its features and usability. We would like to especially acknowledge Brenton Fargnoli, Isaac Klein, York Chen, Sarah Post, Elizabeth Richey, Mark Harris, and Mark Zhang, our early beta users for their invaluable feedback.

Conflicts of Interest

None declared.

References

1. Drazen JM, Shields HM, Loscalzo J. A division of medical communications in an academic medical center's department of medicine. Acad Med 2014 Dec;89(12):1623-1629 [FREE Full text] [doi: 10.1097/ACM.0000000000000472] [Medline: 25186816]
2. Block L, Habicht R, Wu AW, Desai SV, Wang KN, Silva K, et al. In the wake of the 2003 and 2011 duty hours regulations, how do internal medicine interns spend their time? J Gen Intern Med 2013 Aug;28(8):1042-1047 [FREE Full text] [doi: 10.1007/s11606-013-2376-6] [Medline: 23595927]
3. Leonard M. The human factor: the critical importance of effective teamwork and communication in providing safe care. Qual Saf Health Care 2004 Oct 01;13(suppl_1):i85-i90. [doi: 10.1136/qshc.2004.010033]
4. Agarwal R, Sands DZ, Schneider JD. Quantifying the economic impact of communication inefficiencies in U.S. hospitals. J Healthc Manag 2010;55(4):265-81; discussion 281. [Medline: 20812527]
5. Hewett DG, Watson BM, Gallois C, Ward M, Leggett BA. Intergroup communication between hospital doctors: implications for quality of patient care. Soc Sci Med 2009 Dec;69(12):1732-1740. [doi: 10.1016/j.socscimed.2009.09.048] [Medline: 19846246]
6. Wu RC, Lo V, Morra D, Wong BM, Sargeant R, Locke K, et al. The intended and unintended consequences of communication systems on general internal medicine inpatient care delivery: a prospective observational case study of five teaching hospitals. J Am Med Inform Assoc 2013;20(4):766-777 [FREE Full text] [doi: 10.1136/amiajnl-2012-001160] [Medline: 23355461]
7. Schumacher RM, Lowry SZ. National Institute of Standards and Technology.: National Institute of Standards and Technology; 2010. NIST guide to the processes approach for improving the usability of electronic health records URL: https://www.nist.gov/sites/default/files/documents/itl/hit/Guide_Final_Publication_Version.pdf [accessed 2018-03-21] [WebCite Cache ID 6y5ZYRQRU]
8. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530 [FREE Full text] [doi: 10.1197/jamia.M1370] [Medline: 12925543]
9. Ganslandt T, Hackl WO. Findings from the Clinical Information Systems Perspective. Yearb Med Inform 2015 Aug 13;10(1):90-94 [FREE Full text] [doi: 10.15265/IY-2015-037] [Medline: 26293854]
10. Glaser JP. Glaser on Health Care IT: Perspectives from the Decade that Defined Health Care Information Technology (HIMSS Book Series). Boca Raton, Florida, United States: CRC Press; 2016.
11. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
12. Reichheld FF. Harvard Business Review.: Harvard Business Review; 2003 Dec. The One Number You Need to Grow Internet URL: https://hbr.org/2003/12/the-one-number-you-need-to-grow [accessed 2018-03-02] [WebCite Cache ID 6xcX1JaDu]
13. Wilson JR, Corlett N. Evaluation of Human Work, 3rd Edition. Boca Raton, Florida, United States: CRC Press; 2005.

14. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
15. Qualres. Qualitative Research Guidelines Project: Semi-structured Interviews URL: http://www.qualres.org/HomeSemi-3629.html [accessed 2018-03-21] [WebCite Cache ID 6y5aWXdGO]
16. Bullinger HJ. Human Aspects in Computing: Design and Use of Interactive Systems and Work With Terminals?. London: Elsevier Science Ltd; 1991.
17. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
18. Miles MB, Huberman A. Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: SAGE; 1994.
19. Cheung YB. A modified least-squares regression approach to the estimation of risk difference. Am J Epidemiol 2007 Dec 01;166(11):1337-1344. [doi: 10.1093/aje/kwm223] [Medline: 18000021]
20. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004 Apr 01;159(7):702-706. [Medline: 15033648]
21. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
22. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
23. Promoter.io. What's a Good NPS Score? URL: https://www.promoter.io/blog/good-net-promoter-score/ [accessed 2018-03-21] [WebCite Cache ID 6y5bCH1Vn]
24. OpenView Labs. 2016. Net Promoter Score: Vanity or Value Metric? URL: https://labs.openviewpartners.com/net-promoter-score/ [accessed 2018-03-21] [WebCite Cache ID 6y5bJl23o]
25. DirectLineMD. URL: http://www.directlinemd.com/ [accessed 2018-03-02] [WebCite Cache ID 6xcXrYADJ]
26. AgileMD. Improve clinical care URL: https://www.agilemd.com/ [accessed 2018-03-02] [WebCite Cache ID 6xcXumuL9]
27. Rastegar-Mojarad M, Kadolph C, Ye Z, Wall D, Murali N, Lin S. A fuzzy-match search engine for physician directories. JMIR Med Inform 2014 Nov 04;2(2):e30 [FREE Full text] [doi: 10.2196/medinform.3463] [Medline: 25601050]
28. Haroon M, Yasin F, Eckel R, Walker F. Perceptions and attitudes of hospital staff toward paging system and the use of mobile phones. Int J Technol Assess Health Care 2010 Oct;26(4):377-381. [doi: 10.1017/S0266462310001054] [Medline: 20923591]
29. Kummerow Broman K, Kensinger C, Hart H, Mathisen J, Kripalani S. Closing the loop: a process evaluation of inpatient care team communication. BMJ Qual Saf 2017 Jan;26(1):30-32. [doi: 10.1136/bmjqs-2015-004580] [Medline: 26614773]
30. Gadd CS, Penrod LE. Assessing physician attitudes regarding use of an outpatient EMR: a longitudinal, multi-practice study. Proc AMIA Symp 2001:194-198. [Medline: 11825179]

Abbreviations

CSS: cascading style sheets
EHR: electronic health record
LDAP: lightweight directory access protocol
NPS: Net Promoter Score
REDCap: Research Electronic Data Capture
SOAP: simple-object access protocol
SSL: Secure Socket-Layer
SUS: System Usability Scale

Edited by G Eysenbach; submitted 10.10.16; peer-reviewed by Z Ye, C Steele Gray, F Ehrler, R Gunter, A Paglialonga; comments to author 09.01.17; revised version received 20.07.17; accepted 12.02.18; published 16.04.18.

Please cite as:
Morawski K, Monsen C, Takhar S, Landman A
A Novel Information Retrieval Tool to Find Hospital Care Team Members: Development and Usability Study
JMIR Hum Factors 2018;5(2):e14
URL: http://humanfactors.jmir.org/2018/2/e14/
doi:10.2196/humanfactors.6781
PMID:29661752

Review

The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review

Saif Sherif Khairat[1], PhD; Aniesha Dukkipati[1], MSN, RN; Heather Alico Lauria[1], BSN, RN; Thomas Bice[1], MD, MSc; Debbie Travers[1], PhD, MSN, FAEN; Shannon S Carson[1], MD

University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Saif Sherif Khairat, PhD
University of North Carolina at Chapel Hill
428 Carrington Hall
Chapel Hill, NC, 27514
United States
Phone: 1 919843514
Email: saif@unc.edu

Abstract

Background: Intensive Care Units (ICUs) in the United States admit more than 5.7 million people each year. The ICU level of care helps people with life-threatening illness or injuries and involves close, constant attention by a team of specially-trained health care providers. Delay between condition onset and implementation of necessary interventions can dramatically impact the prognosis of patients with life-threatening diagnoses. Evidence supports a connection between information overload and medical errors. A tool that improves display and retrieval of key clinical information has great potential to benefit patient outcomes. The purpose of this review is to synthesize research on the use of visualization dashboards in health care.

Objective: The purpose of conducting this literature review is to synthesize previous research on the use of dashboards visualizing electronic health record information for health care providers. A review of the existing literature on this subject can be used to identify gaps in prior research and to inform further research efforts on this topic. Ultimately, this evidence can be used to guide the development, testing, and implementation of a new solution to optimize the visualization of clinical information, reduce clinician cognitive overload, and improve patient outcomes.

Methods: Articles were included if they addressed the development, testing, implementation, or use of a visualization dashboard solution in a health care setting. An initial search was conducted of literature on dashboards only in the intensive care unit setting, but there were not many articles found that met the inclusion criteria. A secondary follow-up search was conducted to broaden the results to any health care setting. The initial and follow-up searches returned a total of 17 articles that were analyzed for this literature review.

Results: Visualization dashboard solutions decrease time spent on data gathering, difficulty of data gathering process, cognitive load, time to task completion, errors, and improve situation awareness, compliance with evidence-based safety guidelines, usability, and navigation.

Conclusions: Researchers can build on the findings, strengths, and limitations of the work identified in this literature review to bolster development, testing, and implementation of novel visualization dashboard solutions. Due to the relatively few studies conducted in this area, there is plenty of room for researchers to test their solutions and add significantly to the field of knowledge on this subject.

(JMIR Hum Factors 2018;5(2):e22) doi:10.2196/humanfactors.9328

KEYWORDS

intensive care unit; visualization, Dashboard; cognitive load; information overload; usability; user interface design; health information technology; electronic health record

Introduction

State of the Problem—Critical Patient Population

Intensive Care Units (ICUs) in the United States admit more than 5.7 million people each year [1]. The ICU level of care helps people with life-threatening illness or injuries and involves close, constant attention by a team of specially-trained health care providers [2]. ICU patients require frequent assessment and have a greater need for technological and clinical support compared to non-ICU patients [1]. Important metrics in the ICU range from simple vital sign monitoring and laboratory data to mechanical ventilator support, vasoactive medications, and even complete circulatory support, depending on the unique needs of specific patients. Although ICU patients receive care for a wide variety of disease states, the leading causes of death in the ICU are multi-organ system failure, cardiovascular failure, and sepsis [2]. Delay between condition onset and implementation of necessary interventions can dramatically impact the prognosis of a patient with one of these life-threatening diagnoses.

Electronic Health Record Usability

EHR use has increased nationwide; however, the question remains whether EHRs are being used in an effective and efficient way that improves clinical workflow and health outcomes [3,4]. A systematic review and meta-analysis intended to evaluate effects of health information technology in the hospital and ICU on mortality, length of stay, and cost found significant interstudy and intrastudy variability. The study demonstrated that more research is needed with standardized interventions and endpoints to evaluate EHR use and implementation. Currently, no conclusion can be made regarding the effect of health information technology on inpatient and ICU outcomes such as mortality, length of stay, and cost [4].

Information Overload

In 2013, Singh, Spitzmueller, Petersen, Sawhney, and Sittig conducted a cross-sectional study of primary care providers to evaluate predictors of missed test results in the setting of electronic health record (EHR) alerts. Of the nearly 2,600 respondents, 87% perceived the quantity of alerts they received to be excessive, 70% reported receiving more alerts than they could effectively manage, 56% reported that the current EHR notification system made it possible for practitioners to miss test results, and 30% reported having personally missed test results that led to care delays [5]. To address the high volume of metrics used and the time-sensitive nature of responding to changes in a critically ill patient's condition, a tool that improves ICU display and retrieval of key clinical information has great potential to benefit patient outcomes.

Proposed Solution

Visualization is a field of study concerned with the transformation of data to visual representations, where the goal is the effective and efficient cognitive processing of data [6]. Use of visualization techniques in the clinical setting have the potential to improve data display and cognitive processing of data, reducing cognitive overload among clinicians [6]. Information visualization involves the transformation from lower-level data to visual representations of meanings extracted from the data [6]. Extraction is by either a computational process or a human transcription process, the aim of which is to explore data and create new insights [6].

Some guidelines for the development of an information visualization solution include:

- Apply realistic techniques to enhance mapping of data elements to visual objects.
- Minimize user actions to accomplish a goal.
- Provide flexibility in the ways to achieve the same goal.
- Provide functionality to represent additional information.
- Spatially organize the visual layout.
- Consistently apply design choices.
- Place minimal cognitive load on the user.
- Provide users with information on alternatives when several actions are available.
- Remove extraneous or distracting information.
- Consider means to reduce the data set [6].

A dashboard is a data-driven clinical decision support tool capable of querying multiple databases and providing a visual representation of key performance indicators in a single report [7]. The utility of a dashboard comes from its ability to provide a concise overview of key information [7]. Applied to the intensive care unit, a dashboard allows clinicians to quickly identify changes in the patient's condition that require intervention. The clinician can choose to dive deeper into the EHR data or refer to the dashboard at a later point to review changes. Depending on the design of the dashboard, features such as alerts and documentation reminders can help clinicians improve compliance with best practice guidelines and organizational standards [7].

Purpose of this Literature Review

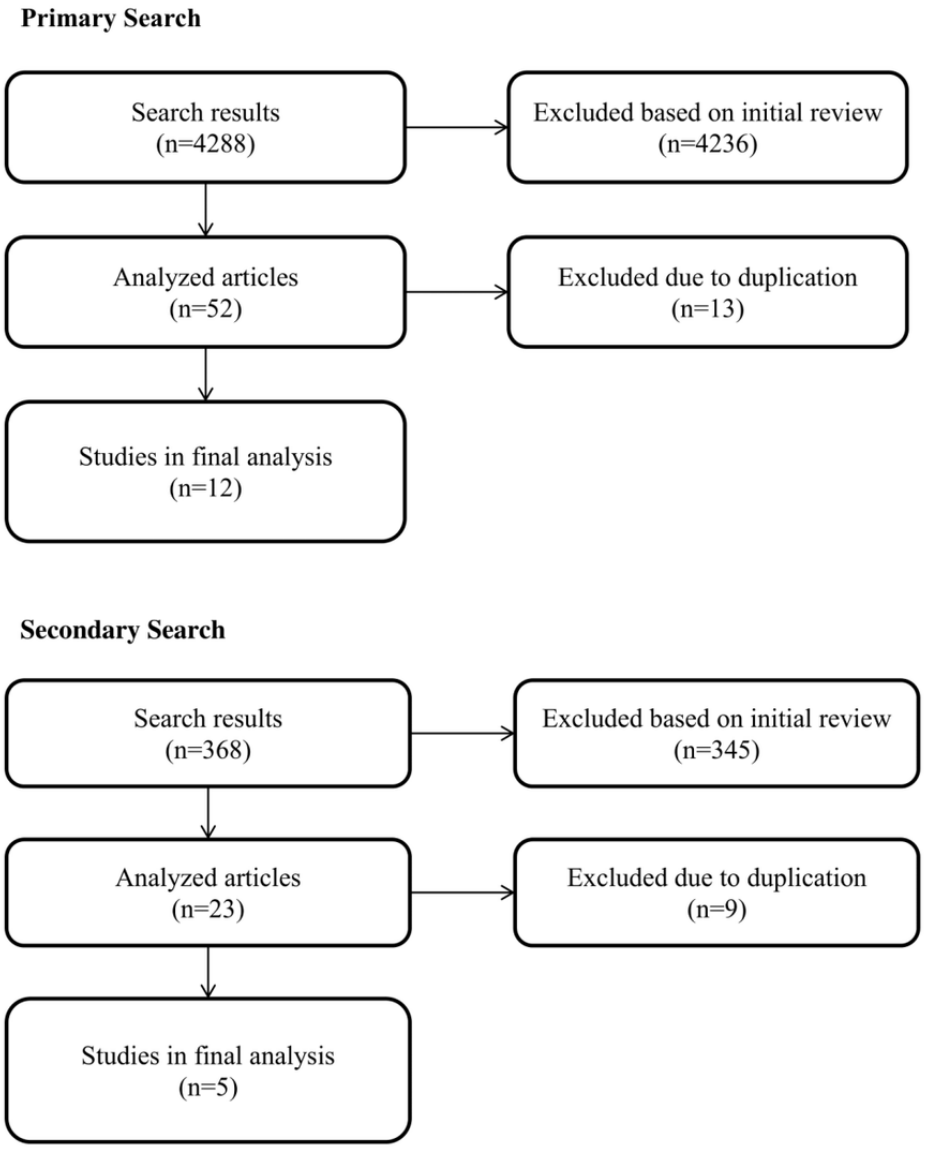
The purpose of conducting this literature review is to present previous research on the use of visualization dashboards to improve efficiency, clinician satisfaction, patient safety and accuracy in the clinical setting. This evidence can be used to guide the development, testing and implementation of new solutions to optimize the visualization of clinical information.

Methods

Inclusion Criteria

Articles were included if they addressed the development, testing, implementation, or use of a visualization dashboard solution in a health care setting. An initial search was conducted of literature on dashboards only in the intensive care unit setting, but there were limited articles found that met the inclusion criteria. A secondary follow-up search was conducted to broaden the results to any health care setting. Ideally, the article would compare outcomes with the novel solution to outcomes prior to or without the novel solution. However, articles were not excluded simply due to lack of a specific comparison. Articles should contain quantitative or qualitative outcomes related to clinician satisfaction, cognitive overload, or patient outcomes. Initially, abstracts were scanned to identify if articles were relevant to the specified research questions.

Figure 1. Literature review process.



Exclusion Criteria

Due to the specificity and novel nature of this topic, no filters were applied to the query. This means that articles were not excluded solely based on type, publication date, or country of origin. However, articles were excluded if there was not an English version of the article available. Articles were excluded if review of the abstract and full text revealed the article did not address at least one of the specified research questions and meet the inclusion criteria.

Databases

Databases selected for this search were PubMed, PMC, CINAHL, and EMBASE, all of which are health sciences journal article databases.

Initial Search Terms—Intensive Care Unit Only

To capture alternative ways of denoting the terms of interest, the query of ("electronic medical record" OR "electronic health record" OR EMR OR EHR) AND ("visualization" OR "dashboard" OR "design" OR "interface") AND ("intensive care" OR "ICU" OR "critical care" OR "CCU") was used for the initial search. Abstracts were screened for relevance to the intended investigation. Articles with relevant abstracts were then read in entirety to further screen the relevance and quality of the data.

Searching the above query into PubMed returned 151 results, 14 of which were analyzed, and eight of which were relevant, quality results for the final analysis. PMC returned 3405 results, 13 of which were analyzed, five of which were excluded due to duplication, and three of which were included in the final analysis. CINAHL returned 86 results, seven of which were analyzed, two of which were excluded due to duplication, and 0 included in the final analysis. EMBASE returned 646 results, 18 of which were analyzed, six of which were excluded due to duplication, and one of which was included in the final analysis. Therefore, a total of 12 articles were obtained from this primary search towards the final analysis [8-19].

Follow-up Search Terms—Any Health Care Setting

Because of the limited number of results obtained with the initial search, a secondary search was completed using the query of visualization AND dashboard. The purpose of the secondary search was to broaden the search to a visualization dashboard solution in any health care setting, as opposed to only the intensive care unit setting. The same databases and process of screening articles were maintained from the initial search process.

Searching the above query into PubMed returned 24 results, seven of which were analyzed, one of which was excluded due to duplication, and four of which were relevant, quality results for the final analysis. PMC returned 311 results, 10 of which were analyzed, five of which were excluded due to duplication, and one of which was included in the final analysis. CINAHL returned three results, two of which were analyzed, none of which were excluded due to duplication, and none of which were included in the final analysis. EMBASE returned 30 results, four of which were analyzed, three of which were excluded due to duplication, and none of which were included in the final analysis. Therefore, a total of five articles were obtained from the secondary search towards the final analysis, Figure 1 [20-24].

Implications of Query Results

The initial and follow-up searches returned a total of 17 articles that were analyzed for this literature review [8-24]. The limited results reflect the novel status of this area of research. Supporting information from information and library science databases will be useful in the analysis steps as much of this work involves development, testing, and implementation of a novel software solution.

Results

The dashboard solutions that were identified in the 17 articles are presented with organization by study findings related to efficiency, quality and safety, accuracy, and user satisfaction. Some solutions were discussed in multiple articles, whereas others were unique to a single article. Table 1 presents the sample size, metrics of interests, results and findings for each study.

Efficiency

The Ambient Warning and Response Evaluation (AWARE) system was tested in two articles included in this literature review [8,9]. AWARE is an ICU-specific patient viewer and monitoring system that was developed at Mayo Clinic [8]. AWARE is a superstructure for existing EHR. The development of this tool was guided by clinicians and based their information needs [8]. Pickering et al (2015) used a step wedge cluster randomization trial to demonstrate a decrease in time spent on pre-round data gathering using the AWARE system [9]. Compared to the existing EHR, AWARE was reported to improve information management (data presentation format and efficiency of data access) and make the task of gathering data for rounds significantly less difficult and mentally demanding [9].

Scripps Clinic and Green Hospital used a rapid-cycle evaluation process to develop the algorithms, alert systems, and interfaces intended to facilitate patient-provider interactions and determination of treatment plans [20]. Brooke's Standardized Usability Tool was used to evaluate usability and two independent appraisers reviewed the think aloud sessions for usability themes [20]. Results pointed to positive results regarding usability and efficiency to identify pertinent components in the patient's plan of care with use of the prototype [20].

Ahmed et al (2011) evaluated a novel .NET-based application by conducting a randomized crossover study [10]. This study demonstrated improved workload (using NASA-task load index), decreased time to task completion, and decreased number of errors of cognition. Additionally, the standard EHR contained a much larger data volume compared with the novel user interface [10]. An image of this patent-pending dashboard is shown in Figure 2.

Koch et al (2013) evaluated nurses' situation awareness and task completion time using an integrated information display compared to traditional displays [11]. Task completion time (response time from seeing the question to submitting the answer) was measured using paper prototypes of both displays [11]. Task completion times were nearly half with integrated displays compared to traditional displays [11]. Figure 3 demonstrates a screenshot of the integration of information displays that was used by Koch et al (2013).

Farri et al (2012) carried out three iterations of planning, risk analysis, design, and evaluation of an EHR prototype. This user interface contained specific functionalities for clinical documents [12]. They used a spiral model for software development and the EHR system user interface framework of the Veterans Affairs computerized patient record system (VistA CPRS) [12].

The researchers used a mixed methods approach to evaluate a sample of eight medical interns as they synthesized EHR clinical documents in four pre-formed clinical scenarios [12]. Despite the non-significant difference in total times to task completion the researchers observed shorter times for two scenarios with the visualization tool. This may suggest that the timesaving benefits may be more evident with certain clinical processes [12].

Dolan et al (2013) used a mixed quantitative and qualitative evaluation process to evaluate their dashboard prototype [21]. The researchers observed the time participants spent using the dashboard before choosing a preferred drug, ease of use, acceptability, decisional conflict, and an open-ended qualitative analysis [21]. Qualitative findings were positive, suggesting potential for informed decision making and patient centered care [21].

Table 1. Study characteristics and results.

Study	Metric of Interest	Sample (n)	Result	Findings
Ahmed et al (2011) [10]	Accuracy, Efficiency	160	Errors Per Provider—Standard: 0.5, AWARE[a]: 2, *P*=.01; Workload-Standard: 38.8, AWARE: 58, *P*<.001; Time-Standard: 145, 125, 129, 112 s, AWARE: 93, 60, 68, 54, *P*<.001, data volume 1008 vs 102.	Less errors per provider, decreased time to task completion for 4 patients, improved workload (NASA-TLX[b]) scores shown after using the visualization tool.
Farri et al (2012) [12]	Accuracy, Efficiency	8	Accuracy: Missing data 2.3 (SD 1.2) with the visualization tool, 6.8 (SD 1.2) without the visualization tool, *P*=.08, accurate inferences 1.3 (SD 0.3) vs. 2.3 (SD 0.3), *P*=.09.	Lower risk missing (unretrieved) patient information with the visualization tool. More accurate inferences. Not statistically significant. Time decreased in two visualization scenarios.
Koch et al (2013) [11]	Accuracy, Efficiency	12	Time-Standard: 42.1 s, Dashboard: 26.0 s, *P*<.001; Accuracy-Standard: 1.8%, Dashboard: 85.3%, *P*<.001.	Nurses had task completion times were nearly half with integrated displays compared to traditional displays.
Clarke et al (2016) [20]	Accuracy, Efficiency, Satisfaction	Mock patients: 15	Analysis of data unavailable.	Discussion of Brooke's Standardized Usability Tool to evaluate usability themes. Examined accuracy and efficiency of Heart Team in identifying pertinent components of patient plan of care.
Faiola et al (2015) [13]	Accuracy, Efficiency, Satisfaction	12	Time-experimental group was faster in answering two questions: [Q3] t(10)=3.11, *P*=.01, r=.70; [Q4] t(10)=3.65, *P*=.004, r=.76; Accuracy-experimental (mean .65, SD .30), control groups (mean .58, SD .36), $\chi 2(1,12)=5.04$, *P*=.03.	Clinical decision-making accuracy was higher when using the visualization dashboard. Faster decision-making on 2/8 questions. Qualitative discussion of potential positive impact of MIVA[c] 2.0
Pickering et al (2015) [9]	Efficiency	Pre: 80, Post: 63	Time on preround data gather; Pre: 12 min, Post: 9 min.	Improved efficiency of information management and data presentation; reduced mental demand.
Dolan et al (2013) [21]	Efficiency, Quality or Safety	25	Mean time interacting with the dashboard=4.6 min. No comparison group.	Interactive clinical decision dashboard are capable of fostering informed patient decision making and patient centered care.
Pageler et al (2014) [15]	Efficiency, Quality or Safety	64	Increased compliance with dressing changes from 87% to 90% (*P*=.003); cap changes 87% to 93% (*P*<.001); port needle changes 69% to 95% *P*<.001); decreased compliance with insertion bundle compliance 67% to 62% *P*=001); 2.6 CLABSIs[d] per 1000 line-days before intervention to 0.7 CLABSIs per 1000 line-days.	Improved compliance with an evidence-based, pediatric-specific catheter care bundle.
Hagland (2010) [17]	Quality/Safety	N/A[e]	No quantifiable data.	Potential to improve patient safety, communication and clinician workflow.

Study	Metric of Interest	Sample (n)	Result	Findings
Shaw et al (2015) [16]	Quality or Safety	450	Time-Pre: 393 min, Post: 202 min, *P*=.05, Quality/Safety-Decreased urinary catheter 16 to 11, *P*=.01	The median time from PICU[f] admission to obtaining treatment consent decreased by 49%. Patients with catheter in place >96 hours decreased from 16 to 11.
Dziadzko et al (2016) [8]	Satisfaction	361	Pre: 15 min, Post: 12 min, *P*=.03.	Less time spent on gathering data using the visualization tool
Bakos et al (2012) [14]	Satisfaction, Quality or Safety	N/A	No quantifiable data.	Increased usage showed clinician satisfaction, benefits for staff per interviews, increased compliance, and decreased adverse events.
Hartzler et al (2015) [23]	Tool Development	Step 1: 6, Step 2: 40	No quantifiable data.	The strategy for tool development was the engagement of healthcare providers to design a user-friendly patient care dashboard.
Badgeley et al (2016) [18]	Tool development	N/A	No quantifiable data.	No data provided; discusses tool development.
Ellsworth et al (2014) [19]	Tool development	23	No quantifiable data.	Large amount of clinical data needed to make clinical decisions; need options for viewing data based on clinical role.
Sebastian et al (2012) [24]	Tool development	N/A	No quantifiable data.	Tool development informed by qualitative data on satisfaction from interviews with neurosurgeons.
Swartz et al (2014) [22]	Tool development	N/A	No quantifiable data.	Survey and structured interview used to create tool. Tool has not been implemented. Better understanding of clinician needs can inform tool development.

[a]AWARE: Ambient Warning and Response Evaluation.

[b]NASA-TLX: NASA Task Load Index.

[c]MIVA: Medical Information Visualization Assistant.

[d]CLABI: central line associated blood stream infection.

[e]N/A: not applicable.

[f]PICU: pediatric intensive care unit.

Medical Information Visualization Assistant, v.2 (MIVA 2.0) is an EHR dashboard technology that uses a visualization engine to deliver multivariate biometric data by transforming it into temporal resolutions [13]. ICU clinicians can use selection menus to control the viewability of data in various time periods to assist with diagnosis and treatment [13]. The usability speed test identified no significant difference in time-on-task between the control group and the experimental group [13]. However, a significant difference was noted in speed with use of MIVA 2.0 [13].

Clinician Satisfaction

Dziadzko et al (2016) studied the before-and-after implementation experience and satisfaction of ICU providers at two hospitals using the AWARE system [8]. Providers agreed that data gathering using the existing EHR system was difficult and time-intensive [8]. In a survey analysis, researchers found that prescribers were significantly more satisfied with the delivery of content and information output with AWARE due to the improvement of the presentation of information [8]. Bakos et al (2012) showed an increased use of the dashboard tool at Virginia Commonwealth University Health System throughout the first year of implementation, demonstrating clinician satisfaction with usage. Interviews further confirmed the benefit and helpfulness of using the tool as staff confirmed its usefulness in their workflow [14].

Quality and Safety

Pageler et al (2013) discuss use of a checklist enhanced by the EHR and a unit-wide dashboard to improve compliance with an evidence-based, pediatric-specific catheter care bundle [15]. The researchers performed a cohort study with historical controls

that included all patients with a central venous catheter at a 24-bed Pediatric ICU (PICU) in an academic children's hospital [15].

Central line associated bloodstream infection (CLABSI) rates decreased after the checklist intervention [15]. Analysis of specific bundle elements demonstrated decreased compliance with insertion bundle documentation. However, there was an increase in compliance with daily documentation of line necessity, dressing changes, cap changes, and port needle changes.

Shaw et al (2015) evaluated a real-time visual display that showed data on presence of consent for treatment, restraint orders, presence of urinary catheters, deep venous thrombosis (DVT) prophylaxis, Braden Q score, and medication reconciliation [16]. An automated EHR querying tool was created to assess compliance with a PICU safety bundle and querying of the EHR for compliance and updating of the dashboard automatically occurred every five minutes [16].

Baseline compliance and duration of noncompliance was established during three time periods: before activation of the dashboard, at one month following activation of the dashboard, and at three months after activation [16]. There was no difference between the three periods in presence of restraint orders, DVT prophylaxis, or development or worsening of pressure ulcers [16]. Between the first and third time periods, the median time from PICU admission to obtaining treatment consent decreased [16]. The number of patients with urinary catheters in place > 96 hours decreased significantly after the intervention [16]. The researchers concluded that a unit-wide dashboard could increase awareness for potential interventions, thereby affecting patient safety in a dynamic manner [16].

Although Bakos et al (2012) speculate that their visualization dashboard will contribute to having zero events of preventable harm to patients, employees and visitors; there is no quantifiable data to support this at this time [14]. Similarly, Hagland (2010) discusses the potential to improve patient safety, communication and clinician workflow using a new clinical dashboard without quantifiable results [17].

Accuracy

Koch et al (2013) used the paper prototypes of their displays to measure situation awareness (accuracy of the participants' answer). Nurses had a higher situation awareness and accuracy when using the integrated display versus the traditional display [11].

To evaluate the accuracy of, MIVA 2.0, Faiola et al (2015) used quantitative clinical decision-making task questions. The clinical decision-making accuracy test identified an overall significant improvement in accuracy of the eight-question test between the experimental versus control groups. Qualitative results were obtained from seven open-ended interview questions, wherein participants acknowledged the potential impact of MIVA 2.0 for reducing cognitive load and enabling more accurate decision-making [13]. Overall, a significant difference was noted in accuracy with use of MIVA 2.0 [13].

Figure 2. "Elements of data are pulled from across the entire electronic medical record and are organized in the systems based manner most commonly encountered in the study's intensive care unit setting." [10].

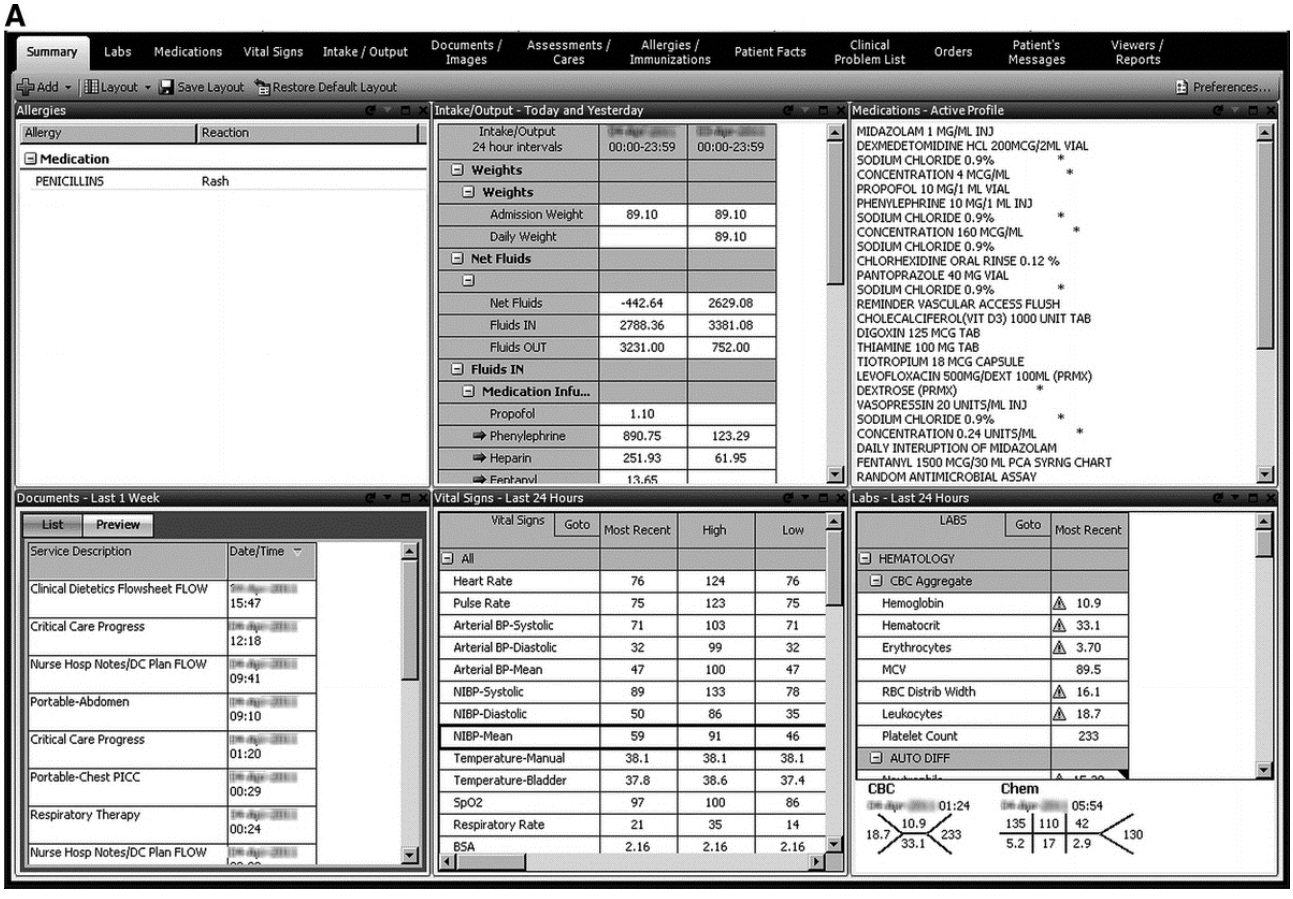
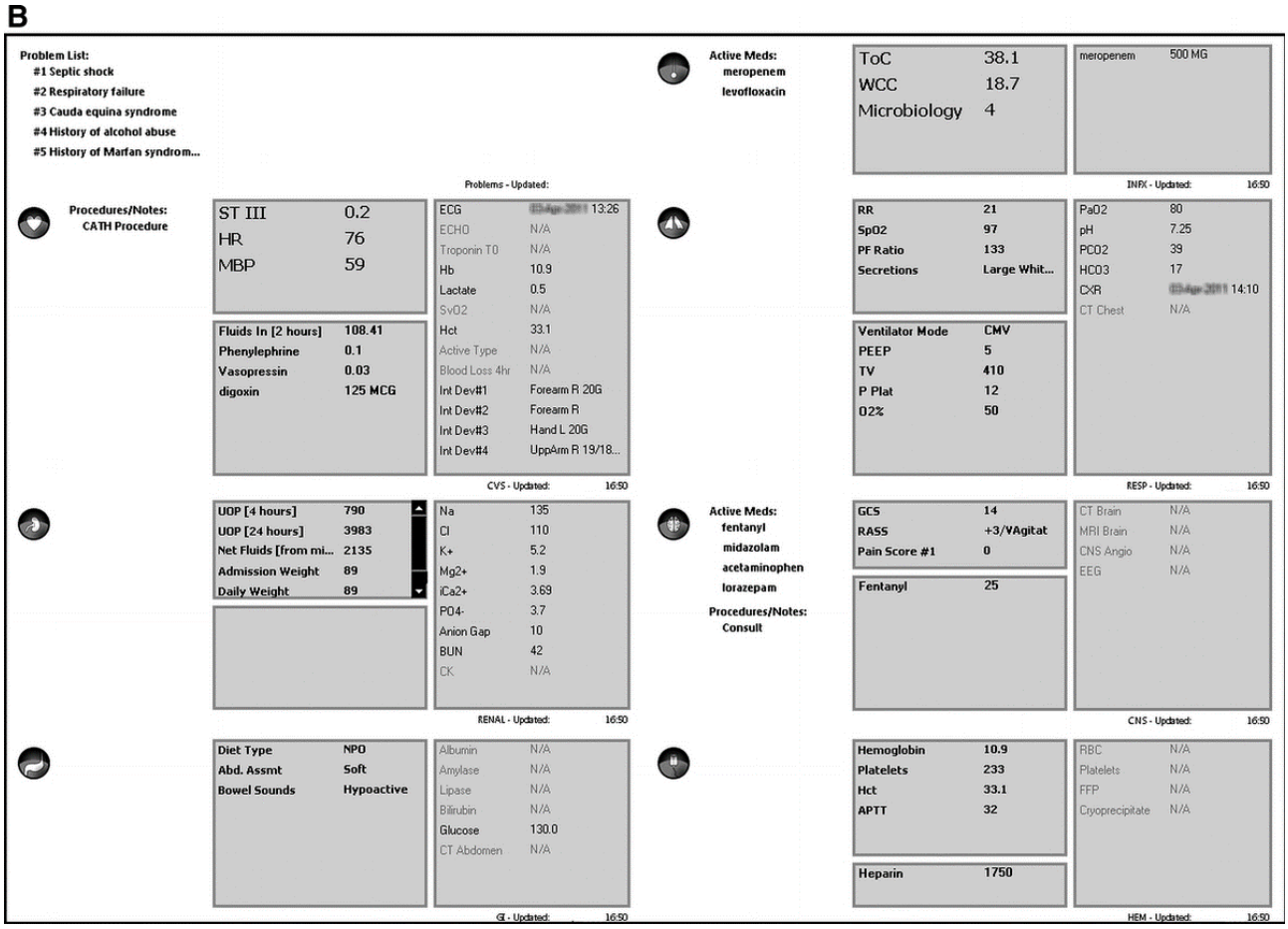


Figure 3. "Elements of data are pulled from across the entire electronic medical record and are organized in the systems based manner most commonly encountered in the study's intensive care unit setting." [10].



Dziadzko et al (2016) surveyed healthcare providers who reported an improvement in the accuracy of decision-making using AWARE, but no quantifiable data is available [8]. Using the .NET based application, Ahmed et al (2011) found that the median number of errors per provider decreased significantly for the novel user interface compared to the standard electronic medical record interface [12].

Farri et al (2012) evaluated the accuracy of using their spiral model software. The resulting differences in unretrieved patient information and accurate inferences were not statistically significant but suggested some improvement with the new information visualization tool [12]. Other observed effects of the tool included more intuitive navigation between patient details and increased effort towards methodical synthesis of clinical documents [12].

Scripps Clinic and Green Hospital demonstrated an improved accuracy of the healthcare provider "Heart Team" in clinical decision-making using 15 mock patients. However, a complete data analysis was not performed [20].

Tool Development

Five articles did not focus on efficiency, quality and safety, accuracy and satisfaction outcomes but discussed their process of tool development. Their findings during visualization tool development are included in the discussion section below [18,19,22-24].

Discussion

Overview

The 17 articles included in this literature review demonstrate how efficiency, quality and safety, clinician satisfaction and accuracy can be improved using a visualization dashboard. These 17 articles share many themes regarding how each dashboard was designed and what user-friendly features are available when using the dashboard [8-24]. These themes are discussed below. With each idea outlined, a discussion of its application to prior visualization dashboard solutions and its implications for future studies follows. Application of these approaches, methods, and features may serve useful in future efforts related to this subject matter. A summary of findings from the articles is depicted in Figures 4 and 5.

Human-Centered Design

A collaborative, human-centered approach informed the creation of several different dashboards. Structured survey and interview were used to inform iterative design and evaluate the final prototype of each dashboard design.

The IView dashboard was developed for use on three ICU's at the Children's Hospital at Pittsburgh and resulted from intensive clinician-IT team-based work and a collaborative relationship with the hospital's clinical IT vendor [13]. Qualitative measures regarding perceived patient safety, clinician workflow, and

physician-nurse communications pointed to positive outcomes in all three categories [13].

Swartz et al (2014) discuss the creation of iNYP, a Java-based service-oriented web application, to meet the specific information needs of emergency medicine clinicians [23]. A combination of survey and structured interview were used to inform the development of this specialty-specific clinical dashboard [22].

Hartzler et al (2015) discuss the use of human-centered design methods to create visual displays of patient reported outcomes [23]. Targeted, iterative design activities were used to inform development of a dashboard that visually displays patient-reported pain and disability outcomes following spine surgery [23]. The Multi-signal Visualization of Physiology (MVP) was developed at the Neuroscience ICU of the National Neuroscience Institute in Singapore to provide a more visual, straightforward, and intuitive diagnosis process [24]. The MVP makes use of a polygram that incorporates live readings of physiological signs and colors to highlight different patient statuses [24].

Interdisciplinary Approach

Nine articles mentioned use of an interdisciplinary approach in developing, testing, and implementing their visualization solution. The benefit of an interdisciplinary approach is that the varied professional perspectives and skills that come with different disciplines are integrated into each step of the process [8,10,13,14,17,19,20,22,23]

Use of an Interactive Prototype

Prototyping is a useful process as it allows developers to strategize product design and obtain feedback from end users without the expansive investment of resources required to make changes in the EHR format [11-13,18,20,21,23]. While there are viable electronic prototyping options available, paper-based prototyping can be a useful, cost-effective solution in the early stages of product design [25]. A mixed evaluation process of quantitative and qualitative measures can be used to direct feedback from end user interaction with the prototype and improve design on subsequent revisions [11-13,18,20,21,23].

Using Open-Source Technology vs Adapting a Third-Party Vendor's Electronic Health Record System

A team with limited resources may not be able to invest financial, temporal, and staff resources into developing a suitable product [26]. Those teams with limited resources may have to wait for a solution to stem from others using open-source technology or for the third-party vendor to provide an option that will be suitable [18,27]. Use of an open-source technology can allow more freedom for the user to develop and share their tool with others than when adjusting an EHR developed by a third-party vendor [18]. Ultimately, each team at a specific organization will decide which route aligns better with their own resources and goals, but the distinct opportunities and risks inherent with each option are important to consider.

Adapting a vendor's EHR system will require continual consultation with the vendor and there may be significant limitations imposed by the contract between the organization and the vendor [27].

Application of Evidence-Based, Clinical Practice Guidelines to the Electronic Health Record

Clinical practice guidelines (CPGs) are intended to improve the quality, consistency, and effectiveness of care by applying evidence-based medicine [28]. A review of physician adherence to clinical practice guidelines suggested that as many as 38% of physicians consider clinical practice guidelines as inconvenient or too difficult to use [28]. Incorporation of clinical practice guidelines into the structure and display of the EHR may help improve convenience of access to practice guidelines and increase use in clinical decision-making [14-16].

Figure 4. "(A) Nurses see an overview of the patient's vital signs, currently administered and scheduled medication, essential ventilation data, and fluid balance. (B) When selecting a medication they see medication compatibility with the other current and scheduled medication, and potential adverse effects." [11].

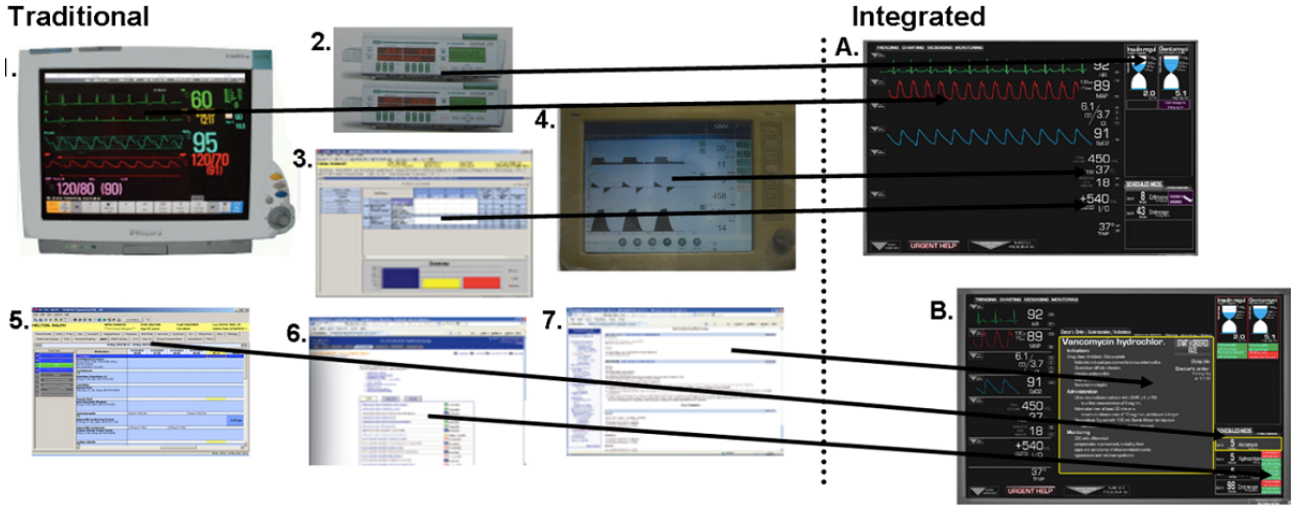
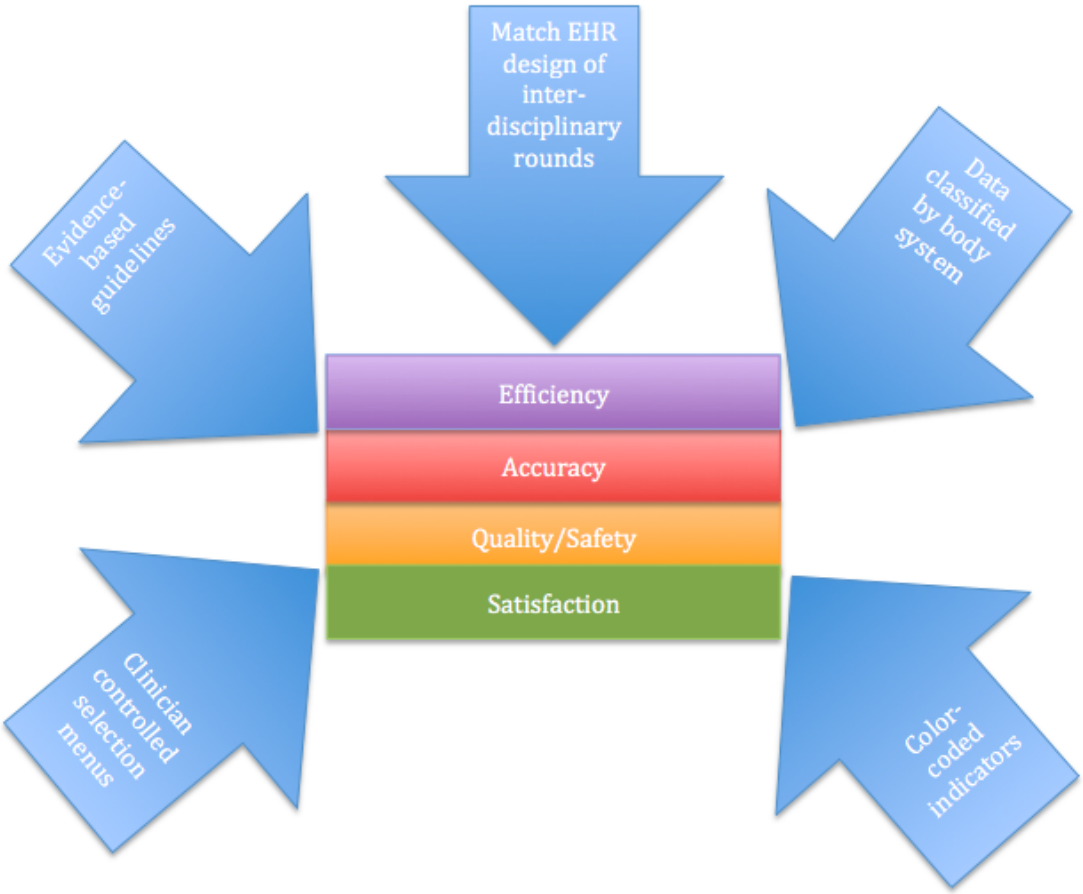


Figure 5. Summary of findings from the literature review. EHR: electronic health record.



Using Open-Source Technology vs Adapting a Third-Party Vendor's Electronic Health Record System

A team with limited resources may not be able to invest financial, temporal, and staff resources into developing a suitable product [26]. Those teams with limited resources may have to wait for a solution to stem from others using open-source technology or for the third-party vendor to provide an option that will be suitable [18,27]. Use of an open-source technology can allow more freedom for the user to develop and share their tool with others than when adjusting an EHR developed by a third-party vendor [18]. Ultimately, each team at a specific organization will decide which route aligns better with their own resources and goals, but the distinct opportunities and risks inherent with each option are important to consider.

Adapting a vendor's EHR system will require continual consultation with the vendor and there may be significant limitations imposed by the contract between the organization and the vendor [27].

Application of Evidence-Based, Clinical Practice Guidelines to the Electronic Health Record

Clinical practice guidelines (CPGs) are intended to improve the quality, consistency, and effectiveness of care by applying evidence-based medicine [28]. A review of physician adherence to clinical practice guidelines suggested that as many as 38% of physicians consider clinical practice guidelines as inconvenient or too difficult to use [28]. Incorporation of clinical practice guidelines into the structure and display of the EHR may help improve convenience of access to practice guidelines and increase use in clinical decision-making [14-16].

Clinician Controlled Selection Menus

Allowing the clinician to adjust the data displayed in alignment with the preference and needs of that individual may further improve clinician satisfaction with the system [13,19]. This capability can also help meet the goal of reducing cognitive overload [13,19]. If clinicians can filter out information that is not pertinent to them, the remaining information will have improved visibility without obstruction from extraneous information [13,19]. The capability to filter information by location, service lines, and specific diagnoses may also serve useful to improve efficiency, accuracy and user satisfaction of clinicians managing many patients [13].

Improved Display of Trends in Physiological Signs

In a setting such as an intensive care unit, the stability of a patient's condition can quickly deteriorate [2]. While clinicians have primary responsibility to assess their patient's condition and intervene appropriately, adding features to the EHR that can assist with this process can expedite these steps; improving efficiency [24]. With the vast array of physiological parameters under continuous monitoring in the ICU setting, improved

display of data trends may improve the clinician's responsiveness in adding or weaning interventions based on the patient's changing condition [24].

Classification of Data by Body System

Classification of data using a body system approach was a common decision for the EHR designs in this literature review [8,10]. By using a body system approach, clinicians can follow a systematic approach to optimizing the patient's holistic health. The design choice of matching the body system approach used by intensive care unit clinicians allows for congruency between the EHR display and cognitive organization of clinical information [8,10].

Applicability of a Visualization Dashboard to Non—Intensive Care Unit Clinical Settings

While this literature review focused primarily on the application of a visualization dashboard to the intensive care unit setting, the same intervention could have benefit in other clinical settings as well [12,14,18,20-23]. The emergency department could be well suited for this intervention as the EHR could then assist in alerting clinicians to new results and a change in the patient's clinical status that modifies the plan of care [22]. Step-down units and inpatient floors may not have the same extent of clinical data as the intensive care unit setting but clinicians may still find benefit from features related to improved display of clinical information. Once a visualization dashboard is successfully implemented in the ICU setting, the dashboard can be modified, tested, and implemented in non-ICU clinical settings; working towards similar goals [18].

Strengths and Limitations of Solutions

This literature review includes information on several visualization dashboards that have been tested with positive results from quantitative and qualitative analysis. These positive results support the potential benefits of a visualization dashboard solution to clinical practice environments. Limitations were noted in the following areas:

- The interpretation of what a visualization dashboard solution entails varied widely among the researchers of the different studies included.
- Many of the visualization dashboard solutions were evaluated with a solely qualitative approach, rather than with a quantitative or mixed methods approach.
- Some articles included details about the design, implementation, and evaluation processes, but did not include full detail on the data obtained.
- Some studies used a simulated setting in lieu of a live clinical setting, which means that results may differ when the solution is applied to a live clinical setting.
- Most studies tested a single solution in a single implementation setting, which limits the generalizability of the findings to other solutions and other implementation settings.

Future Direction

Researchers can build on the findings, strengths, and limitations of the work identified in this literature review to bolster development, testing, and implementation of a novel visualization dashboard solution. Due to the relatively few studies conducted in this area, there is plenty of room for researchers to test their solutions and add significant information to the field of knowledge on this subject. An effective solution in this area can drive process improvement and improved patient outcomes for not only the initial setting of implementation, but also to any further clinical units and organizations that adopt the intervention.

Conclusions

Overall, successful visualization dashboards utilized an interdisciplinary approach to develop a human-centered design. Dashboards were flexible and could be adjusted to the users' preferences as well as organized based on body system, color-coded and adapted for clinician team rounding. These features are important due to the variety in patient population and the diverse way that clinicians interpret information. Utilizing these common themes to develop visualization tools for patient care has shown to improve efficiency, quality or safety, clinician satisfaction and accuracy in a variety of patient settings.

This section synthesizes the major findings of the 17 articles [8-24]. As discussed, visualization tools have the potential to impact accuracy, efficiency, user satisfaction, quality or safety of care in the ICU and other settings. Numerous factors such as clinician-controlled displays, organization by body system, an interdisciplinary design team and using open-source technology can result in successful implementation of a visualization dashboard. The findings, strengths, and limitations discussed in this section can drive future research efforts on visualization dashboard solutions.

Design Recommendation based on Clinician Needs

Information needs varied based on patient population and clinical role. Key findings regarding clinician needs for the solution included: the application of evidence-based, clinical practice guidelines; clinician-controlled selection menus; the use of color-coded visual indicators; classification of data by body system and matching of EHR design to the process of interdisciplinary rounds. As demonstrated in the results section of this paper, the combination of the above components can allow for user-friendly dashboard designs that have the potential to impact accuracy, efficiency, user satisfaction and quality and safety of care.

Conflicts of Interest

None declared.

References

1. Society OCCM. Critical Care. 2017. Critical care statistics URL: http://www.sccm.org/Communications/Pages/CriticalCareStats.aspx [accessed 2018-05-09] [WebCite Cache ID 6zI8W9sJ3]
2. U.S. National Library of Medicine. 2016. Critical care URL: https://medlineplus.gov/criticalcare.html [WebCite Cache ID 6zI8ZIMac]
3. Charles D, Gabriel M. Searcy. In: Adoption of electronic health record systems among U.S. non-federal acute care hospitals. ONC Data Brief. Washington DC: Office of the National Coordinator for Health Information Technology; 2015:2008-2014.
4. Thompson G, O'Horo JC, Pickering BW, Herasevich V. Impact of the Electronic Medical Record on Mortality, Length of Stay, and Cost in the Hospital and ICU: A Systematic Review and Metaanalysis. Crit Care Med 2015 Jun;43(6):1276-1282. [doi: 10.1097/CCM.0000000000000948] [Medline: 25756413]
5. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA Intern Med 2013 Apr 22;173(8):702-704 [FREE Full text] [doi: 10.1001/2013.jamainternmed.61] [Medline: 23460235]
6. Engelbrecht L, Botha A, Alberts R. Designing the Visualization of Information. Int. J. Image Grap 2015 Apr;15(02):1540005. [doi: 10.1142/S0219467815400057]
7. Wilbanks BA, Langford PA. A review of dashboards for data analytics in nursing. Comput Inform Nurs 2014 Nov;32(11):545-549. [doi: 10.1097/CIN.0000000000000106] [Medline: 25310224]
8. Dziadzko MA, Herasevich V, Sen A, Pickering BW, Knight AA, Moreno FP. User perception and experience of the introduction of a novel critical care patient viewer in the ICU setting. Int J Med Inform 2016 Apr;88:86-91. [doi: 10.1016/j.ijmedinf.2016.01.011] [Medline: 26878767]
9. Pickering BW, Dong Y, Ahmed A, Giri J, Kilickaya O, Gupta A, et al. The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial. Int J Med Inform 2015 May;84(5):299-307. [doi: 10.1016/j.ijmedinf.2015.01.017] [Medline: 25683227]
10. Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. Crit Care Med 2011 Jul;39(7):1626-1634. [doi: 10.1097/CCM.0b013e31821858a0] [Medline: 21478739]
11. Koch SH, Weir C, Westenskow D, Gondan M, Agutter J, Haar M, et al. Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: A prospective randomized controlled study. Int J Med Inform 2013 Aug;82(8):665-675. [doi: 10.1016/j.ijmedinf.2012.10.002] [Medline: 23357614]
12. Farri O, Rahman A, Monsen KA, Zhang R, Pakhomov SV, Pieczkiewicz DS, et al. Impact of a prototype visualization tool for new information in EHR clinical documents. Applied Clinical Informatics 2012;3(3):404-418. [doi: 10.4338/ACI-2012-05-RA-0017] [Medline: 23646087]
13. Faiola A, Srinivas P, Duke J. Supporting Clinical Cognition: A Human-Centered Approach to a Novel ICU Information Visualization Dashboard. AMIA Annu Symp Proc 2015;2015:560-569 [FREE Full text] [Medline: 26958190]
14. Bakos KK, Zimmermann D, Moriconi D. Implementing the Clinical Dashboard at VCUHS. NI 2012 (2012) 2012;2012:11 [FREE Full text] [Medline: 24199038]
15. Pageler NM, Longhurst CA, Wood M, Cornfield DN, Suermondt J, Sharek PJ, et al. Use of electronic medical record-enhanced checklist and electronic dashboard to decrease CLABSIs. Pediatrics 2014 Mar;133(3):e738-e746 [FREE Full text] [doi: 10.1542/peds.2013-2249] [Medline: 24567021]
16. Shaw SJ, Jacobs B, Stockwell DC, Futterman C, Spaeder MC. Effect of a Real-Time Pediatric ICU Safety Bundle Dashboard on Quality Improvement Measures. Jt Comm J Qual Patient Saf 2015 Sep;41(9):414-420. [Medline: 26289236]
17. Hagland M. First Place. Children's Hospital of Pittsburgh. Improving patient care through data availability in the ICU. Healthc Inform 2010 Mar;27(3):20-1, 23, 26 passim. [Medline: 20344925]
18. Badgeley MA, Shameer K, Glicksberg BS, Tomlinson MS, Levin MA, McCormick PJ, et al. EHDViz: clinical dashboard development using open-source technologies. BMJ Open 2016 Mar 24;6(3):e010579 [FREE Full text] [doi: 10.1136/bmjopen-2015-010579] [Medline: 27013597]
19. Ellsworth MA, Lang TR, Pickering BW, Herasevich V. Clinical data needs in the neonatal intensive care unit electronic medical record. BMC Med Inform Decis Mak 2014 Oct 24;14:92 [FREE Full text] [doi: 10.1186/1472-6947-14-92] [Medline: 25341847]
20. Clarke S, Wilson ML, Terhaar M. Using Clinical Decision Support and Dashboard Technology to Improve Heart Team Efficiency and Accuracy in a Transcatheter Aortic Valve Implantation (TAVI) Program. Stud Health Technol Inform 2016;225:98-102. [Medline: 27332170]
21. Dolan J, Veazie P, Russ A. Development and initial evaluation of a treatment decision dashboard. BMC Med Inform Decis Mak 2013 Apr 21;13:51 [FREE Full text] [doi: 10.1186/1472-6947-13-51] [Medline: 23601912]
22. Swartz JL, Cimino JJ, Fred MR, Green RA, Vawdrey DK. Designing a clinical dashboard to fill information gaps in the emergency department. AMIA Annu Symp Proc 2014;2014:1098-1104 [FREE Full text] [Medline: 25954420]
23. Hartzler AL, Chaudhuri S, Fey BC, Flum DR, Lavallee D. Integrating Patient-Reported Outcomes into Spine Surgical Care through Visual Dashboards: Lessons Learned from Human-Centered Design. EGEMS (Wash DC) 2015;3(2):1133 [FREE Full text] [doi: 10.13063/2327-9214.1133] [Medline: 25988187]

24. Sebastian K, Sari V, Loy LY, Zhang F, Zhang Z, Feng M. Multi-signal Visualization of Physiology (MVP): a novel visualization dashboard for physiological monitoring of Traumatic Brain Injury patients. In: Conf Proc IEEE Eng Med Biol Soc. Multi-signal Visualization of Physiology (MVP): A novel visualization dashboard for physiological monitoring of traumatic brain injury patients. Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE. 2012; 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Nov 1-4, 1990; Philadelphia, Pennsylvania p. 2000-2003. [doi: 10.1109/EMBC.2012.6346349]
25. Sutcliffe A, de Bruijn O, Thew S, Buchan I, Jarvis P, McNaught J, et al. Developing visualization-based decision support tools for epidemiology. Information Visualization 2012 May 24;13(1):3-17. [doi: 10.1177/1473871612445832]
26. Ghazisaeidi M, Safdari R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of Performance Dashboards in Healthcare Sector: Key Practical Issues. Acta Inform Med 2015 Oct;23(5):317-321 [FREE Full text] [doi: 10.5455/aim.2015.23.317-321] [Medline: 26635442]
27. Marsolo K. Informatics and operations--let's get integrated. J Am Med Inform Assoc 2013 Jan 01;20(1):122-124 [FREE Full text] [doi: 10.1136/amiajnl-2012-001194] [Medline: 22940670]
28. Armstrong JJ, Goldfarb AM, Instrum RS, MacDermid JC. Improvement evident but still necessary in clinical practice guideline quality: a systematic review. Journal of Clinical Epidemiology 2017 Jan;81:13-21. [doi: 10.1016/j.jclinepi.2016.08.005] [Medline: 27565978]

Abbreviations

AWARE: Ambient Warning and Response Evaluation
CLABI: central line associated blood stream infection.
CPGs: Clinical Practice Guidelines
DVT: Deep Vein Thrombosis
EHR: Electronic Health Record
ICU: Intensive Care Unit
MIVA: Medical Information Visualization Assistant.
MVP: Multi-signal Visualization of Physiology
NASA-TLX: NASA Task Load Index.
PICU: pediatric intensive care unit.

Edited by G Eysenbach; submitted 01.11.17; peer-reviewed by B Jin, G Lim; comments to author 27.12.17; revised version received 01.03.18; accepted 22.04.18; published 31.05.18.

Please cite as:
Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS
The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review
JMIR Hum Factors 2018;5(2):e22
URL: http://humanfactors.jmir.org/2018/2/e22/
doi:10.2196/humanfactors.9328
PMID:

Original Paper

A Web-Based Treatment Decision Support Tool for Patients With Advanced Knee Arthritis: Evaluation of User Interface and Content Design

Hua Zheng[1], PhD; Milagros C Rosal[1], PhD; Wenjun Li[1], PhD; Amy Borg[1], MPH, MEd; Wenyun Yang[1], MS; David C Ayers[1], MD; Patricia D Franklin[1], MBA, MPH, MD

Department of Orthopedics and Physical Rehabilitation, University of Massachusetts Medical School, Worcester, MA, United States

Corresponding Author:
Patricia D Franklin, MBA, MPH, MD
Department of Orthopedics and Physical Rehabilitation
University of Massachusetts Medical School
55 Lake Ave North
Worcester, MA, 01655
United States
Phone: 1 (508) 856 5748
Email: patricia.franklin@umassmed.edu

Abstract

Background: Data-driven surgical decisions will ensure proper use and timing of surgical care. We developed a Web-based patient-centered treatment decision and assessment tool to guide treatment decisions among patients with advanced knee osteoarthritis who are considering total knee replacement surgery.

Objective: The aim of this study was to examine user experience and acceptance of the Web-based treatment decision support tool among older adults.

Methods: User-centered formative and summative evaluations were conducted for the tool. A sample of 28 patients who were considering total knee replacement participated in the study. Participants' responses to the user interface design, the clarity of information, as well as usefulness, satisfaction, and acceptance of the tool were collected through qualitative (ie, individual patient interviews) and quantitative (ie, standardized Computer System Usability Questionnaire) methods.

Results: Participants were older adults with a mean age of 63 (SD 11) years. Three-quarters of them had no technical questions using the tool. User interface design recommendations included larger fonts, bigger buttons, less colors, simpler navigation without extra "next page" click, less mouse movement, and clearer illustrations with simple graphs. Color-coded bar charts and outcome-specific graphs with positive action were easiest for them to understand the outcomes data. Questionnaire data revealed high satisfaction with the tool usefulness and interface quality, and also showed ease of use of the tool, regardless of age or educational status.

Conclusions: We evaluated the usability of a patient-centered decision support tool designed for advanced knee arthritis patients to facilitate their knee osteoarthritis treatment decision making. The lessons learned can inform other decision support tools to improve interface and content design for older patients' use.

(JMIR Hum Factors 2018;5(2):e17) doi:10.2196/humanfactors.8568

KEYWORDS

usability evaluation; patient decision support; knee osteoarthritis; total knee replacement; outcome prediction

Introduction

Arthritis, with its most common form osteoarthritis (OA), affects 50% of all adults older than 65 years of age and is the most common chronic condition and cause of disability in the United States [1]. When knee OA pain and disability advances, total knee replacement (TKR) surgery can effectively eliminate pain and improve function. Total knee replacement is now the number one most common procedure among hospital discharges [2]. Knowledge about the medical and surgical treatments and associated outcomes is critical for patient decision making. Patients who delay the procedure until late in the symptom course may have less optimal results [3].

Hudak and team [4] evaluated the reasons that prevented some elderly OA patients from considering total joint surgery. Ongoing deliberation of the surgical option mainly resulted in a deferral of the treatment decision. The barriers to limiting surgical decision making include inaccurate estimation of symptom level for surgical candidates and lack of information to discuss with their physicians. Our recent review confirms the lack of published research on shared decision making and patient decision aids in orthopedic surgery, with no evidence about the use of patient decision support tools for knee OA patients considering TKR [5]. Therefore, patient decision support tools for the advanced knee arthritis population are needed to help them understand their individual OA symptom severity, provide evidence-based benefits and risks, and aid communication with their physicians to guide treatment decisions.

The average knee OA patient who chooses surgery is 66 years of age [2], and the user interface both for data entry and data output must be designed to facilitate ease of use among aging adults and minimize potential barriers. The objective of this study was to examine user experience and acceptance of a Web-based treatment decision support tool for advanced knee arthritis patients who are considering TKR surgery. Our results may inform user interface and outcome presentation design for other decision support tools for older adults with diverse health conditions.

Methods

Tool Development

The tool's user interface was designed by a multidisciplinary team including an orthopedic specialist, a researcher with expertise in health literacy, a computer scientist, and a biostatistician. The team focused on developing a user interface design that would be simple to operate by older adults with functional limitations such as vision decline and diminished motor skills. To facilitate use among low literacy individuals, the tool used white background and dark text, one-question-per-page display, big font and simple layout, and plain language within eighth-grade literacy reading level.

Briefly, the tool prompts patients to respond to 20 questions related to demographics, overall health, knee pain and function, medical comorbidities, and expectations one year after surgery. Using data entered by the patient, the tool estimates likely individual patient-level improvement in post-TKR pain relief and physical function according to patient characteristics and current health attributes. These estimates are then translated into metrics meaningful to patients (ie, pain relief at rest, pain relief when walking, and ability to walk five blocks at a year after surgery). These metrics are easily understood by patients and can be used to facilitate communication between patients and surgeons and thus support TKR decision making.

Patient Recruitment

The study sample was recruited from the UMass Memorial Health Care Arthritis and Total Joint Center. All patients aged 21 years of age and older seeking knee OA care at the Arthritis and Total Joint Center were eligible. Patients with acute knee injuries or who were not fluent in English were excluded.

A study recruiter screened all new pre-TKR and post-TKR patients during the study months. After confirming eligibility, a study coordinator contacted each potential participant by telephone to describe the study and invite him/her to participate. If the patient was willing to take part in the study, the study coordinator scheduled an interview before or after the next doctor's appointment, and mailed a fact sheet, a consent form, and a HIPAA authorization form to the patient for signature. At the interview, the study coordinator answered any questions and gave a copy of the consent form to the patient in case he/she did not bring the signed one. Patient participants received a stipend of US $10 for parking at the end of the interview. The study was approved by Institutional Review Board for the protection of human subjects.

Usability Evaluation Procedures

User-centered formative and summative evaluations were employed for the tool usability testing [6]. The first phase goal, the formative evaluation, was to improve the tool design through participants' response to preliminary ideas of design. This was accomplished through a first round of evaluation interviews. The second phase, designed as the summative evaluation, was to assess the clarity of outcome information as well as usefulness, satisfaction, and acceptance of the tool through interviews and questionnaires. This was accomplished through a second round of evaluation interviews. Methods used in each round are described subsequently.

Round 1

Round 1 was performed based on the iterative evaluation process; the tool was adjusted after each subround of interviews and was then reassessed in the next subround. To avoid bias, different participants were recruited in each subround. Round 1 interviews started with a survey of patient demographics and computer abilities. Participants were asked to use the Web-based tool on the computer and encouraged to think aloud their immediate feelings as they completed each survey page and task. The think-aloud method was used to verbalize users' thoughts, feelings, and opinions while interacting with the system. Thinking aloud slows the thought process and increases mindfulness, which is very helpful for capturing a wide range of cognitive activities. During the use of the tool, the participants were asked questions about tool design. The questions were structured with predetermined topics, such as wording, layout, color, button and overall utility, as well as with open-ended comments. The overall duration of the interview was up to 30 minutes. The process was administered by a study coordinator with expertise in patient interviews.

Round 2

Round 2 was a summative evaluation to conduct an overall assessment of the near-final version of the tool. Round 2 patients were asked to report their demographics and computer abilities at the beginning of the interview. They then completed the survey questions of the tool with no interruption, followed by an interview about their opinions about the presentation of outcome information. Five types of presentations were provided: text summaries, bar graphs, word clouds, smiling faces, and staged walking people. Interview items assessed the format that

was liked best / liked least, the ease of understanding, and the helpfulness for decision making. Finally, participants completed a standard Computer System Usability Questionnaire (CSUQ) [7], a standardized assessment to measure user satisfaction with computer system usability and four customized research questions on technical difficulty and tool format preference (Web vs paper). The interview process took up to 30 minutes. The number of participants enrolled in round 2 was based on group size-specified guides for quantitative usability studies [8,9].

Data Collection Tools

Patient participants completed study procedures in a quiet room adjacent to the Arthritis and Total Joint Center. A desktop computer allowed access to the Web-based tool with survey questions and outcome data display. Screen recorder software, Camtasia Studio 6, was used to captured user's operations on the computer screen, such as cursor movement, mouse clicking, and keyboard input. A digital voice recorder taped the comments and discussion during the process. Participants' gender, age, education level, and computer use were asked on a one-page demographics and computer ability survey. An interview guide was developed by the study team based on user-centered formative and summative evaluations. A trained interviewer administered patient interview process and a usability specialist acted as primary observer.

Data Analysis

Patient demographics and computer ability data were analyzed descriptively. Means and proportions were used to describe the characteristics of the study sample. Qualitative analysis summarized findings from the interview and observation data into several topics, enumerated the patients' needs and preference of design of the tool. Quantitative data included time spent on each page and in total on the use of the tool and usability scores which assessed usefulness, satisfaction, and acceptance of the tool.

Results

Participant Characteristics

For round 1, 11 patients were contacted and 8 (73%) participated in the study. For round 2, 20 patients were contacted and all (100%) participated. Participant characteristics are shown in Table 1. For all 28 patients, the mean age was 63 (SD 11) years and 12 (43%) were older than 65 years; 21 (75%) were female, 26 (93%) had at least a high school education level, 15 (54%) used a computer every day, and 9 (32%) rarely or never used a computer. Eight of nine participants with low computer use were 65 years of age or older.

Round 1 Findings

Eight participants were involved in round 1 and three subrounds were conducted during the iterative design process. The findings are categorized by information clarity and interface design tasks.

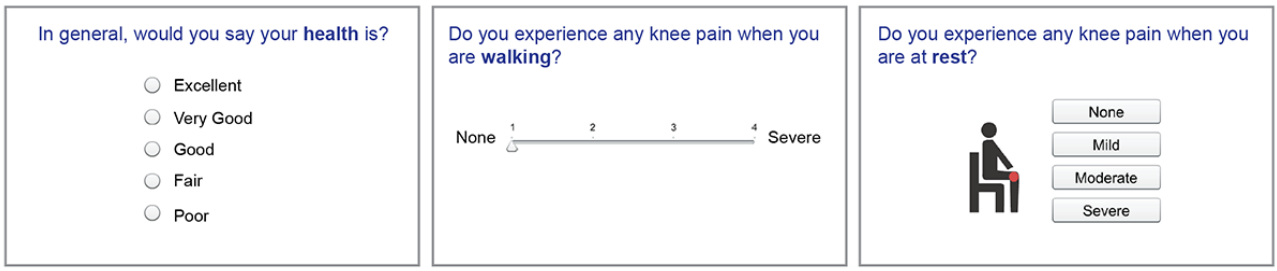
Information Clarity

Most participants had no difficulty in understanding the survey questions. They felt that the language was simple and the wording was easy to understand. The only unclear item was the use of the word "knee scope" in a question about prior surgery. After we changed "scope" to "arthroscopic surgery (a scope inserted by a doctor into your knee)," participants agreed that the presentation was clear. Some participants suggested asking questions for knee pain and physical activity on a good day, a moderate day, and a bad day.

Table 1. Participant characteristics (N=28).

Patient factors	Round 1, n (%) (n=8)	Round 2, n (%) (n=20)
Gender		
Female	7 (88)	14 (70)
Male	1 (12)	6 (30)
Age (years)		
<65	6 (75)	10 (50)
≥65	2 (25)	10 (50)
Education		
Less than high school	0 (0)	2 (10)
Attended or graduated from high school/GED	3 (37)	7 (35)
Attended or graduated from college	5 (63)	11 (55)
Computer use		
Every day	5 (63)	10 (50)
Once a week	0 (0)	3 (15)
Less than once a week but more than once a month	0 (0)	1 (5)
Rarely or never	3 (37)	6 (30)

Figure 1. Questionnaire recording patients' buttons preferences.



(a) radio buttons (b) horizontal slider (c) text buttons

Interface Design

Layout

The questions were organized as one per screen, which was reported as clear and easy to read. Arial was used as the main font and all participants liked it. Font size was modified from 15-18 points to 18-20 points because 15 points was too small. Patients reported that some screens looked similar; for example, knee pain when walking and at rest. Based on this input, we modified the screens to include relevant images, such as someone walking or sitting, and highlighted important words in bold to clarify the difference in question focus.

Buttons

Radio buttons, horizontal sliders, and text buttons (Figure 1) were tested by participants to indicate their answer to a Likert-scale question. Many participants suggested using bigger radio buttons to click. Most participants, especially those who use a computer infrequently, advised against slider buttons because they found that "clicking and dragging" is hard to operate. Text buttons were thought better and easier to click. A "Next Question" button was initially put on each screen with a "Previous Question" button. Some participants expressed the need for a reminder to click on the "Next Question" button. Therefore, we used "automatic jump" to next screen by selecting an answer instead of an extra click on a "Next Question" button. Participants preferred this automatic function.

Colors

We used dark text on a white background for tool screens. No patient had problems with this style. One participant with glaucoma said questions were easy to read. The topic of each question was highlighted on the top of the screen with white text on a dark background; colored backgrounds were initially used to represent different categories of topics; for example, orange for demographics, blue for knee condition, but some participants did not like the colors. To simplify, the final version only used blue for topic background. One participant suggested color-coding the answer to a question in red, yellow, or green when it is relevant, such as red for severe pain and green for no pain, to highlight different selections.

Images

We added images to some of the questions for better comprehension. Most of participants reported that images made questions visually distinct. Numerous participants preferred images with a real person as compared to a "fake" person and one participant did not like cartoon images. A computer-savvy participant felt little attention was given to images compared to words.

Round 2 Findings

Based on the problems identified in round 1 usability testing, we revised the design of the tool. Twenty patients participated in round 2 and tested the enhanced version. Round 2 focused on the testing of the presentation of the outcomes and usefulness, satisfaction, and acceptance of the tool.

Preferred Outcome Presentations

Five outcome presentation formats were shown to participants: text summaries, bar graphs, word clouds, smiling faces, and staged walking people (Multimedia Appendix 1). Participants could choose up to two preferred presentation formats. They easily distinguished which result format appealed to them more, and had clear reactions to different presentations. Bar graphs and staged walking people were preferred overall.

Time Spent on Tool and Each Screen

A total of 19 of 20 Camtasia data records from round 2 were captured; one record was not saved due to an operational error. Four participants seemed unfamiliar with computer use from their records of mouse clicking and keyboard entry. Two of them were older patients who were not able to use the computer themselves and asked the interviewer to operate the mouse for them. Considering the remaining 15 participants, the total time spent on the tool varied between 2 and 4 minutes, and the mean time spent on each screen was 9.7 (SD 4.2) seconds (Figure 2).

Two questions took participants a longer time than others to answer: (1) What is your height and weight? (to answer this question, a participant had to move the cursor to three different boxes and type in their answers), and (2) Have you been told by a health care provider what your knee condition is due to (one of the following)? The distribution of the time spent on each question is in Figure 3.

Figure 2. Distribution of the time spent on each page of the tool.

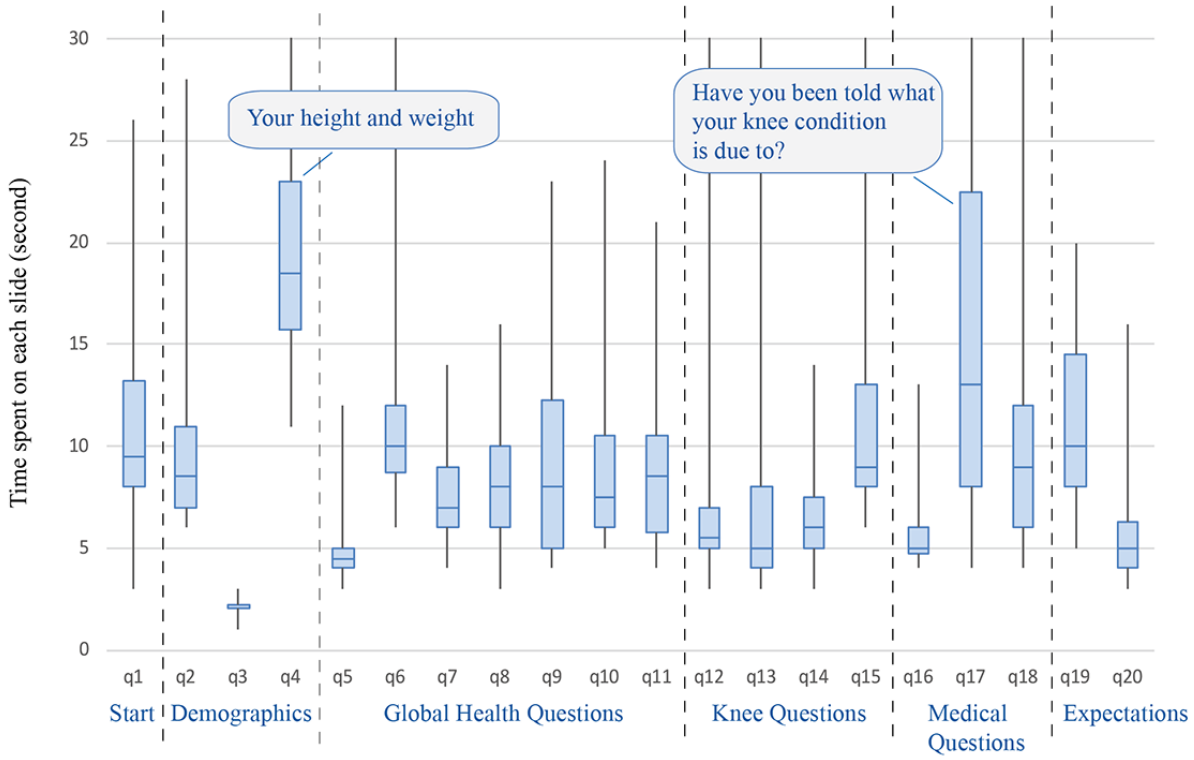
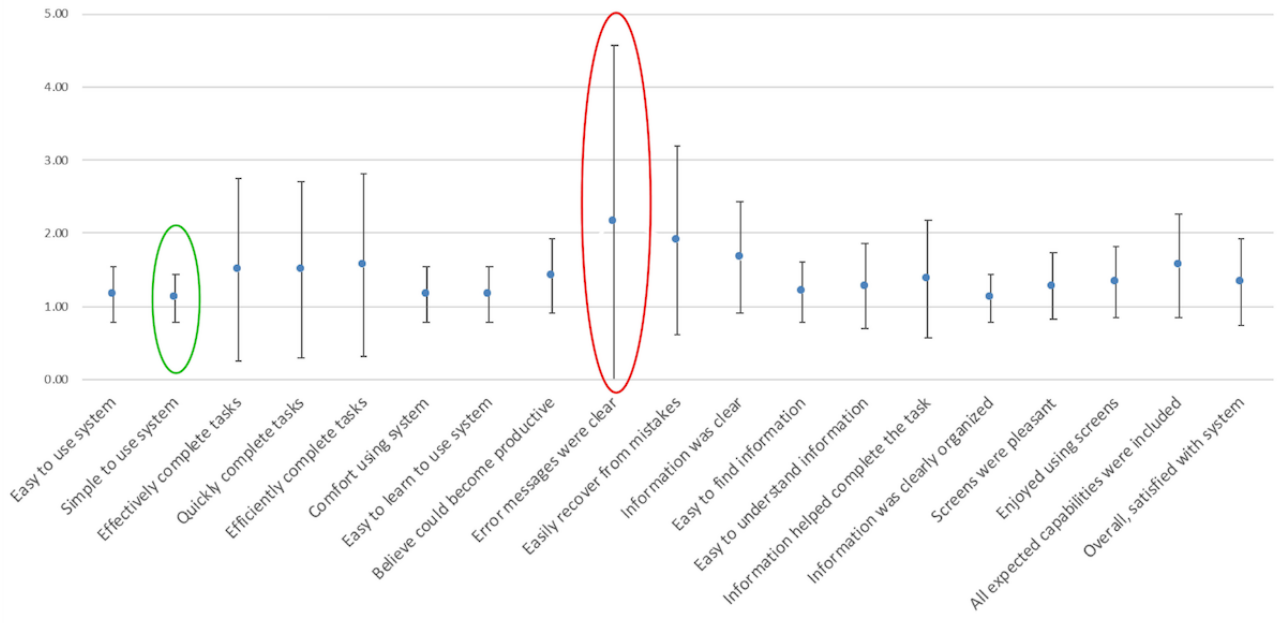


Figure 3. Computer System Usability Questionnaire item scores.



User Satisfaction Scores

A total of 19 of 20 participants in round 2 completed the usability evaluation survey. The questionnaire included 19 items with a Likert scale ranging from 1 (“strongly agree”) to 7 (“strongly disagree”) to measure user satisfaction. Low scores are better than high scores. The mean CSUQ scores on the four items about system usefulness, information quality, interface quality, and overall were 1.32 (SD 0.55), 1.44 (SD 0.58), 1.39 (SD 0.37), and 1.37 (SD 0.41), respectively. The mean and SD for each item can be found in Figure 3. Satisfaction was greatest for simplicity of use (mean 1.11, SD 0.32) and lowest for error message (mean 2.17, SD 2.40). The responses to four customized research questions revealed that 74% (14/19) of the participants had no technical questions using the tool, and 84% (16/19) could use the tool without instructions. In all, 68%

(13/19) of the participants preferred a computer version compared with a paper version, and 26% (5/19) thought either was fine. Finally, 89% (17/19) of participants reported that they would recommend this tool to a friend.

Discussion

Role of Web-Based Treatment Decision Tools

Computerized decision support tools are a new approach to treatment planning [10]. In contrast to patient education systems, decision tools provide information about a recommended treatment plan, including potential benefits and harms, and provide a foundation for patients to make a data-driven decision between two or more treatment options. Web-based interactive tools can facilitate this process by accessing online health information and helping patients get informed treatment options before communicating with doctors [11-13].

Lessons Learned on User Interface

Through usability evaluation of a Web-based patient-centered decision support tool for advanced knee OA patients, we learned the preferences of older OA patients to inform tool design. Textbox 1 is the summary of patient preferences.

Aging adults are an important and understudied group for evaluation of Web survey usability and outcome data presentation. Their needs and concerns may differ from those of other age groups due to the natural changes associated with the aging process. The literature on Web accessibility for older users describes aging-related functional limitations, such as vision decline, motor skill diminishment, and cognitive decline [14,15].

The guidelines for accessible content include large print, simple language, and easy navigation. Our findings are consistent with prior research. For example, the participants liked larger fonts, larger text-filled buttons, fewer colors, simpler navigation without extra “next page” click, less mouse movement, and clear illustrations with simple graphs. Advanced functionality can cause usability difficulties for older adults. For example, horizontal sliders are a common element in Web design, but none of the participants liked them and they reported “am not able to manipulate” or “have difficulty figuring out how to do.”

Textbox 1. Summary of patient preferences.

Interface

- ***Text***
  1. Sans serif font, such as Arial
  2. Big font size of 18 points or more
  3. Highlighting important words
- ***Buttons***
  1. Big text buttons; no slider bars
  2. Automatic jump to next page by selecting an answer instead of an extra click on “Next” button
  3. Avoid operations that need more mouse movement
- ***Colours***
  1. White background and dark text
  2. Fewer unnecessary colors
- ***Images***
  1. Simple images for illustration
  2. Eliminating distracting images

Information

- ***Clarity***
  1. Plain language instead of medical terms
  2. Short description for necessary medical terms

Predictive Outcome Measures

- ***Preferences***
  1. Clear and easy to understand, such as bar charts
  2. Outcome-specific with positive action, such as walking people for arthritis patients

Medical terminology is usually a significant obstacle for patients. Past research has revealed that participants experience difficulties understanding jargon, especially medical terminology [14]. The team's health literacy specialist advised us to avoid medical jargon and improve explanations during the tool development. Thus, most participants reported no problems understanding the information presented. For medical terms that are difficult to simplify, such as arthroscopic surgery, we used both the term and a short description (ie, a scope inserted by a doctor) and learned that this way effectively conveyed the medical information.

The results also revealed that an easy-to-use system is more important than a comprehensive user manual. Most of the participants preferred the computer version over a paper survey. The most recent Pew reports released in 2018 showed that 66% of American adults ages 65 and older use the internet, and 73% of people aged 50 to 64 years and almost one-half of people aged 65 and older own a smartphone [16,17]. Most aging baby boomers will use computerized and mobile tools in the future, so we anticipate growing ease of use.

Lessons Learned on Presenting Outcomes

Presenting likely outcomes of surgical procedures can provide new insights to patients about possible benefits and risks. Tailored estimates of the likely benefits of TKR surgery based on specific patient profiles are feasible using current computing technologies. However, the manner of presentation of predicted outcomes affects how patients understand the value of a treatment and may influence patients' decisions [18,19]. For example, among key TKR outcome research publications, outcomes were expressed as global health assessment scores such as the Short Form Health Survey, the Veterans RAND 12-item Health Survey, or Patient-Reported Outcomes Measurement Information System Global Health survey, or knee-specific pain and function scores such as the the Western Ontario and McMaster Universities Osteoarthritis Index or Knee injury and Osteoarthritis Outcome Score survey. Global and knee outcome metrics are useful to clinicians and researchers, but do not convey to patients likely achievable and meaningful outcomes. In this study, we translated outcome measures into meaningful metrics to patients, such as pain-free walking, or home and community activity levels. The metrics were easily understood by patients, which can facilitate informed communications between patients and surgeons.

In addition, outcome data can be illustrated in different ways and patient comprehension may differ when information is presented using different words or displays to communicate [20-22]. To explore patients' preferences on presentation of outcome data, we evaluated five different presentation formats. It was hypothesized that older adults might prefer text more than numbers, but only a few people chose text display. Color-coded bar charts made more sense to them and were reported to be "clearer" and "easy to understand." These findings are consistent with prior studies that found that bar charts were most commonly preferred and least often found difficult to interpret [20,23]. We were surprised that participants liked the staged walking people display, a combination of graphs and numbers. Walking people graphs were thought more user-friendly and easy to understand, and suggestive of their primary goal—greater activity. We learned that older adults understand and accept outcome-specific graphs with positive action to present data.

Study Limitations

Study limitations include a relatively small sample. However, user interface evaluation research has reported that 31% of usability problems can be identified with a single user [24], and more than 80% of usability problems can be identified with a sample of five users [25,26]. Thus, it is likely that the size of our sample was sufficient to identify most interface design problems. Culturally and linguistically diverse patients were not considered in patient selection in this study. We plan to test the tool in a broad and diverse national sample in the future.

Conclusion

We evaluated the usability of a patient-centered decision support tool designed for advanced knee arthritis patients to facilitate their surgical treatment decision making. Patient participants showed high satisfaction and acceptance of the usefulness and interface quality of this easy, simple tool and selected acceptable data presentation formats for understanding of predictive outcomes after surgery. We expect to collect more data in future studies to verify the qualitative and quantitative findings. Our experience with the tool user interface and outcome presentation design for knee OA patients can inform the design for other chronic conditions within elderly populations.

Acknowledgments

This research was supported by the Robert Wood Johnson Foundation 63839, AHRQ P50HS018910, and PCORI 1507-31714. The authors would like to thank Betsy Costello for her contribution to participant enrollment and Dr Sylvie Puig for her editorial assistance with this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Votes for outcome presentations.

[PDF File (Adobe PDF File), 96KB - humanfactors_v5i2e17_app1.pdf]

References

1. United States Bone and Joint Initiative. The Burden of Musculoskeletal Diseases in the United States: Prevalence, Societal and Economic Cost. 3rd edition. Rosemont, IL: United States Bone and Joint Initiative; 2015.
2. Williams SN, Wolford ML, Bercovitz A. National Center for Health Statistics, Centers for Disease Control and Prevention. 2015 Aug. NCHS Data Brief, no 210: Hospitalization for total knee replacement among inpatients aged 45 and over: United States, 2000-2010 URL: https://www.cdc.gov/nchs/products/databriefs/db210.htm [accessed 2018-04-11] [WebCite Cache ID 6ybTgI1Mb]
3. Fortin PR, Penrod JR, Clarke AE, St-Pierre Y, Joseph L, Bélisle P, et al. Timing of total joint replacement affects clinical outcomes among patients with osteoarthritis of the hip or knee. Arthritis Rheum 2002 Dec;46(12):3327-3330 [FREE Full text] [doi: 10.1002/art.10631] [Medline: 12483739]
4. Hudak PL, Clark JP, Hawker GA, Coyte PC, Mahomed NN, Kreder HJ, et al. "You're perfect for the procedure! Why don't you want it?" Elderly arthritis patients' unwillingness to consider total joint arthroplasty surgery: a qualitative study. Med Decis Making 2002;22(3):272-278. [doi: 10.1177/0272989X0202200315] [Medline: 12058784]
5. Adam JA, Khaw F, Thomson RG, Gregg PJ, Llewellyn-Thomas HA. Patient decision aids in joint replacement surgery: a literature review and an opinion survey of consultant orthopaedic surgeons. Ann R Coll Surg Engl 2008 Apr;90(3):198-207 [FREE Full text] [doi: 10.1308/003588408X285748] [Medline: 18430333]
6. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usabil Stud 2009;5(1):34-45.
7. Scholtz J. Usability Evaluation. 2004. URL: http://notification.etisalat.com.eg/etisalat/templates.backup.16082011/582/Usability%2520Evaluation_rev1%5B1%5D.pdf [accessed 2017-07-26] [WebCite Cache ID 6sFkAVkwo]
8. Nielsen J. Nielson Norman Group. 2006 Jun 26. Quantitative studies: how many users to test? URL: https://www.nngroup.com/articles/quantitative-studies-how-many-users/ [accessed 2017-07-26] [WebCite Cache ID 6sFjuQvpX]
9. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Inter 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
10. Taylor CA, Draney MT, Ku JP, Parker D, Steele BN, Wang K, et al. Predictive medicine: computational techniques in therapeutic decision-making. Comput Aided Surg 1999;4(5):231-247. [doi: 10.1002/(SICI)1097-0150(1999)4:5<231::AID-IGS1>3.0.CO;2-Z] [Medline: 10581521]
11. O'Connor AM, Rostom A, Fiset V, Tetroe J, Entwistle V, Llewellyn-Thomas H, et al. Decision aids for patients facing health treatment or screening decisions: systematic review. BMJ 1999 Sep 18;319(7212):731-734 [FREE Full text] [Medline: 10487995]
12. Elwyn G, Kreuwel I, Durand MA, Sivell S, Joseph-Williams N, Evans R, et al. How to develop web-based decision support interventions for patients: a process map. Patient Educ Couns 2011 Feb;82(2):260-265. [doi: 10.1016/j.pec.2010.04.034] [Medline: 20627644]
13. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
14. Kutner M, Greenberg E, Jin Y, Paulsen C. The Health Literacy of America's Adults Results From the 2003 National Assessment of Adult Literacy. Washington, DC: National Center for Education Statistics; 2006. URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2017-07-26] [WebCite Cache ID 6sFjbLNSq]
15. Arch A. Web Accessibility for Older Users: A Literature Review. 2008 May 14. URL: https://www.w3.org/TR/wai-age-literature/ [accessed 2017-07-26] [WebCite Cache ID 6sFjN3YuY]
16. Pew Research Center. 2018 Feb 05. Internet/Broadband fact sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2017-07-26] [WebCite Cache ID 6sFiit3wT]
17. Pew Research Center. 2018 Feb 05. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-04-11] [WebCite Cache ID 6ybTntsO6]
18. Moxey A, O'Connell D, McGettigan P, Henry D. Describing treatment effects to patients. J Gen Intern Med 2003 Nov;18(11):948-959. [doi: 10.1046/j.1525-1497.2003.20928.x]
19. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. JAMA 2004 May 19;291(19):2359-2366. [doi: 10.1001/jama.291.19.2359] [Medline: 15150208]
20. Carling CL, Kristoffersen DT, Flottorp S, Fretheim A, Oxman AD, Schünemann HJ, et al. The effect of alternative graphical displays used to present the benefits of antibiotics for sore throat on decisions about whether to seek treatment: a randomized trial. PLoS Med 2009 Aug;6(8):e1000140 [FREE Full text] [doi: 10.1371/journal.pmed.1000140] [Medline: 19707579]
21. Feldman-Stewart D, Kocovski N, McConnell BA, Brundage MD, Mackillop WJ. Perception of quantitative information for treatment decisions. Med Decis Making 2000;20(2):228-238. [doi: 10.1177/0272989X0002000208] [Medline: 10772360]
22. Feldman-Stewart D, Brundage MD, Zotov V. Further insight into the perception of quantitative information: judgments of gist in treatment decisions. Med Decis Making 2007;27(1):34-43. [doi: 10.1177/0272989X06297101] [Medline: 17237451]

23. Hildon Z, Allwood D, Black N. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. Int J Qual Health Care 2012 Feb;24(1):55-64. [doi: 10.1093/intqhc/mzr072] [Medline: 22166856]
24. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Apr 24 Presented at: CHI'93 Conference; Apr 24-29, 1993; Amsterdam, The Netherlands p. 206-213. [doi: 10.1145/169059.169166]
25. Nielsen, Jakob. Nielson Norman Group. 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 2017-07-26] [WebCite Cache ID 6sFj6EY7O]
26. Turner C, Lewist J, Nielsen J. Determining usability test sample size. In: Int Enc Ergonom Hum Fact. 2006 Presented at: CHI '93; Apr 24-29, 1993; Amsterdam, The Netherlands p. 3084-3088. [doi: 10.1201/9780849375477.ch597]

Abbreviations

CSUQ: Computer System Usability Questionnaire
OA: osteoarthritis
TKR: total knee replacement

Edited by G Eysenbach; submitted 26.07.17; peer-reviewed by G Peterson, H Umapathy; comments to author 23.11.17; revised version received 01.03.18; accepted 19.03.18; published 30.04.18.

Please cite as:
Zheng H, Rosal MC, Li W, Borg A, Yang W, Ayers DC, Franklin PD
A Web-Based Treatment Decision Support Tool for Patients With Advanced Knee Arthritis: Evaluation of User Interface and Content Design
JMIR Hum Factors 2018;5(2):e17
URL: http://humanfactors.jmir.org/2018/2/e17/
doi:10.2196/humanfactors.8568
PMID:29712620

Original Paper

A Moderated e-Forum for Adults With Cardiovascular Disease: Usability Study

Rika Tanaka[1], PhD; Anita Banerjee[1]; Jelena Surikova[1], BA; Jacqueline Tracey[1], MSc; Ada Payne[2], PhD; Heather Ross[3,4], MD, MHSc, FRCPC, FACC; Robert Nolan[1,3], PhD

[1]Cardiac eHealth and Behavioural Cardiology Research Unit, Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

[2]Models of Care, Clinical Programs and Quality Initiatives, Cancer Care Ontario, Toronto, ON, Canada

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

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

Corresponding Author:
Rika Tanaka, PhD
Cardiac eHealth and Behavioural Cardiology Research Unit
Peter Munk Cardiac Centre
University Health Network
585 University Avenue, 6NU-618
Toronto, ON, M5G 2N2
Canada
Phone: 1 416 340 4800 ext 6400
Email: rika.tanaka@uhnresearch.ca

Abstract

Background: Self-care behaviors are commonly prescribed to manage both cardiovascular disease and hypertension to reduce modifiable risk factors and improve quality of life. Nevertheless, long-term adherence to self-care recommendations for cardiac patients has been problematic. In cardiac patients, moderated online forums have been found to be particularly useful in supporting maintenance of heart-healthy diet and fewer hospital visits. As such, we developed the e-Forum, a Web-based moderated forum designed to promote continued user engagement and long-term self-care adherence.

Objective: The objective of this study was to assess the usability of the user interface for the newly designed e-Forum. In addition to overall user satisfaction, we obtained feedback from our target users on the key features of this newly developed interface.

Methods: An iterative design tested the usability of the e-Forum. On the basis of the user feedback, adjustments were made to the design of our e-Forum, and these changes were then tested in the succeeding group. Participants were recruited from the Heart Function Clinic at the Peter Munk Cardiac Center, University Health Network. After consenting to participate in our study, patients were asked to complete a set of goal-oriented tasks and a feedback interview for the e-Forum. A content analysis of the transcripts from the set of goal-oriented tasks and feedback interviews identified several themes, including general feedback and comments regarding 3 key areas of the e-Forum: layout, navigation, and content.

Results: Overall, 13 cardiac patients (aged 32-81 years) participated in 3 rounds of testing. Participants across all 3 rounds were highly satisfied with our e-Forum and indicated that they would find such a forum useful in managing their health. Expressions of overall satisfaction with the e-Forum and positive comments regarding layout increased between the initial and the final round. As improvements were made to the e-Forum based on participant feedback, potential barriers, negative comments related to the content, and the number of navigation errors decreased between rounds 1 and 3.

Conclusions: We found evidence to support the usability of the user interface for our e-Forum. These results indicate that the e-Forum will likely be a successful tool to support an online community of cardiac patients in their efforts to sustain long-term lifestyle behavior change.

(JMIR Hum Factors 2018;5(2):e20) doi:10.2196/humanfactors.8820

KEYWORDS

support groups; cardiovascular disease; qualitative research

Introduction

Overview

According to the American Heart Association, cardiovascular disease (CVD) accounted for approximately 1 in every 3 deaths in the United States in 2013 [1]. Self-care behaviors (eg, maintaining a healthy diet, regular exercise, and medication adherence) are recommended to manage both CVD and hypertension to reduce modifiable risk factors and improve quality of life [2]. Nevertheless, long-term adherence to self-care recommendations for cardiac patients has been problematic [3]. In an effort to reduce risk for CVD and improve quality of life for patients, our research team developed a Web-based lifestyle counseling platform for cardiac patients (eg, those diagnosed with hypertension or heart failure, HF) to promote adherence to self-care recommendations [4-9].

On the basis of evidence from our program of research, our team created the Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure (CHF-CePPORT; ClinicalTrials.gov: NCT01864369) [10]. Although CHF-CePPORT provides a 12-month comprehensive e-counseling program for self-care behavior change in patients with HF, long-term adherence to Web-based lifestyle counseling programs can be difficult to sustain. For example, dropout rates in Web-based interventions can range up to 62%, and failure to participate in the e-based interventions is 28% over 9 months [11]. These findings indicate that such programs may benefit from supplementary features that facilitate long-term patient engagement and adherence. To address this issue, we developed the e-Forum to supplement CHF-CePPORT by supporting the establishment of an online community that aims to promote continued user engagement and long-term self-care adherence. Our aim was to tailor the design and functional features of the e-Forum to meet the needs of patients with cardiovascular conditions such as HF, who are likely to be older and to present with lower computer literacy. In keeping with guidelines suggested from previous research [12-14], this study assessed the usability of this e-Forum to determine whether cardiac patients could use this program as intended.

Web-Based Moderated Forums

The use of online social networks is an important method for facilitating information sharing as well as providing and receiving support among patients and health care professionals [15,16]. Online communities offer patients access to both emotional support and information about disease management that are not always available or easily accessible [17]. Online moderated forums are online communities that are monitored by professionals or trained peers who (1) facilitate user engagement in the online forum, (2) ensure the accuracy of information discussed by users, and (3) check for safety and appropriateness of posted messages (eg, monitoring for language suggesting self-harm or aggressive or offensive language). Patients demonstrate a preference for this type of intervention over and above conventional e-pages that only present information [18,19]. Such forums have been found to help a diverse array of patients, including those suffering from obesity [17] and ovarian cancer [20]; they offer users a resource to manage the complexities of their illnesses by promoting and supporting healthy self-care strategies [21]. In cardiac patients, such online communities have been found to be particularly useful in supporting maintenance of heart-healthy diet and fewer visits to the hospital [22].

We designed our e-Forum to provide a reliable and accessible interface to foster an online community for patients enrolled in our CHF-CePPORT program. From a functional perspective, our e-Forum was developed to allow users to submit *posts*, including comments or questions regarding their efforts to begin or maintain therapeutic changes in self-care behaviors. The e-Forum was organized such that posts may be submitted under highlighted topics, including “Active Living,” “Eating Healthy,” “Smoke-free Living,” and “Getting Motivated” (see Figures 1-3 to view the final version of the e-Forum). The e-Forum was designed to then send submitted posts to a moderator, who was trained to review posts for accuracy and appropriateness of content and patient safety before they were made accessible and viewable to the other members of the online community. In addition, the e-Forum was designed to allow members of our team to host live or taped presentations on select topics related to self-care adherence and quality of life. The original prototype of the e-Forum also featured large buttons, bright and inviting colors, and large font sizes to increase usability for our older target patient population.

Usability Assessment

Although there is preliminary evidence that the use of online forums may be an effective mode of intervention to enhance education and therapeutic support for participants, it is unclear which features enable users to interact with such forums more effectively [23-25]. Therefore, we undertook a usability study to assess our high-quality, user-centered interface designed to maximize the engagement with the e-Forum [26]. Specifically, our usability study was conducted to determine whether the target users (ie, cardiac patients) could use the e-Forum as intended. Usability studies have been found to improve the design of several other Web-based programs. For example, Stinson et al conducted a usability study to improve their Web-based self-management program for adolescents with arthritis and their parents [27]. In the first of 2 rounds of usability testing, adolescents with arthritis and their parents reported that the labels used in the medication home page were ambiguous, resulting in navigation difficulties in that portion of the program. On the basis of this feedback, the team revised the labeling, and this issue was not reported in the second round of the usability testing. A usability study of a Web-based self-management program for patients diagnosed with chronic obstructive pulmonary disease also found this type of assessment to be helpful in improving the design of their program [13]. The CHF-CePPORT program prototype also underwent a usability study [14]. During this study, navigation issues were identified and resolved before its launch as part of a randomized controlled trial [14]. Together, these studies suggest that users can provide practical feedback to help identify problems with functionalities that may have otherwise been overlooked.

Figure 1. Final version of e-Forum home page. On the basis of user feedback, the final version of the e-Forum home page included a button on the upper left corner that allows users to add a post to the forum without first accessing a topic. It also included a search function as well as a scroll button on the upper right corner.

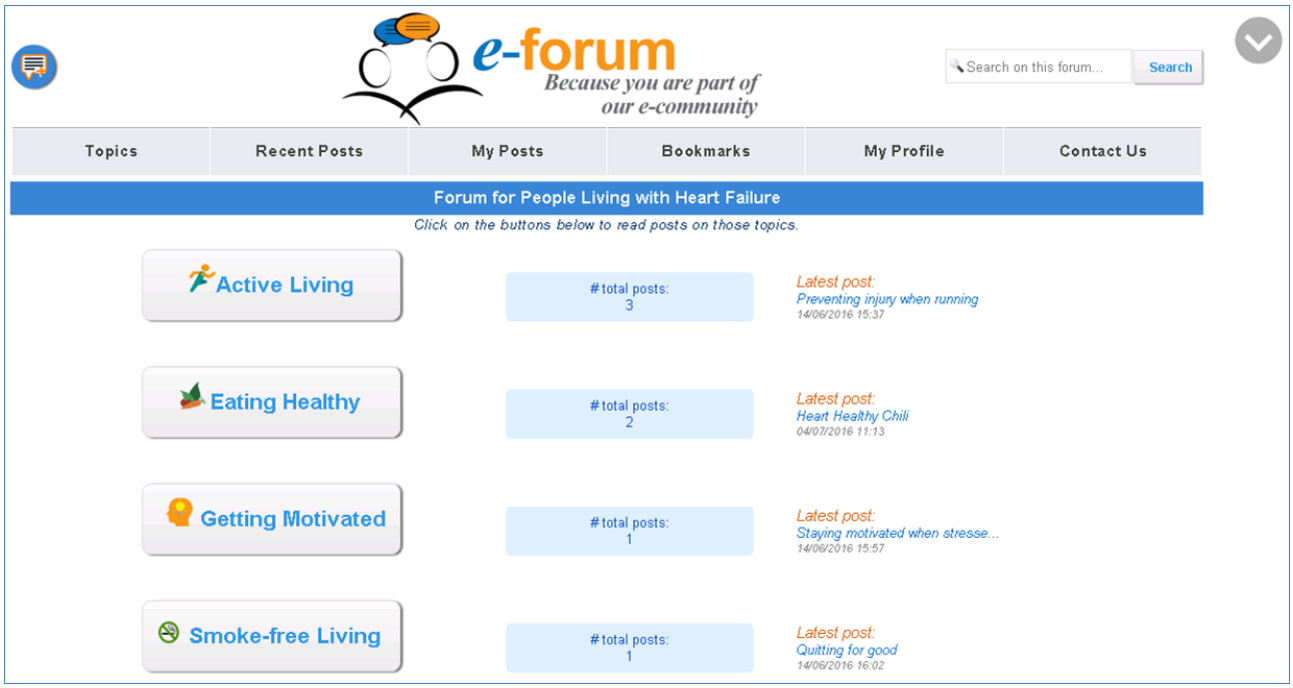


Figure 2. Final version of post thread. On the basis of user feedback, the final version of the post thread included a breadcrumb feature on the upper left corner that allows users to see what topic and thread they are reading. Buttons on this page were also redesigned to consistently include both icons and verbal descriptions of the button functions.

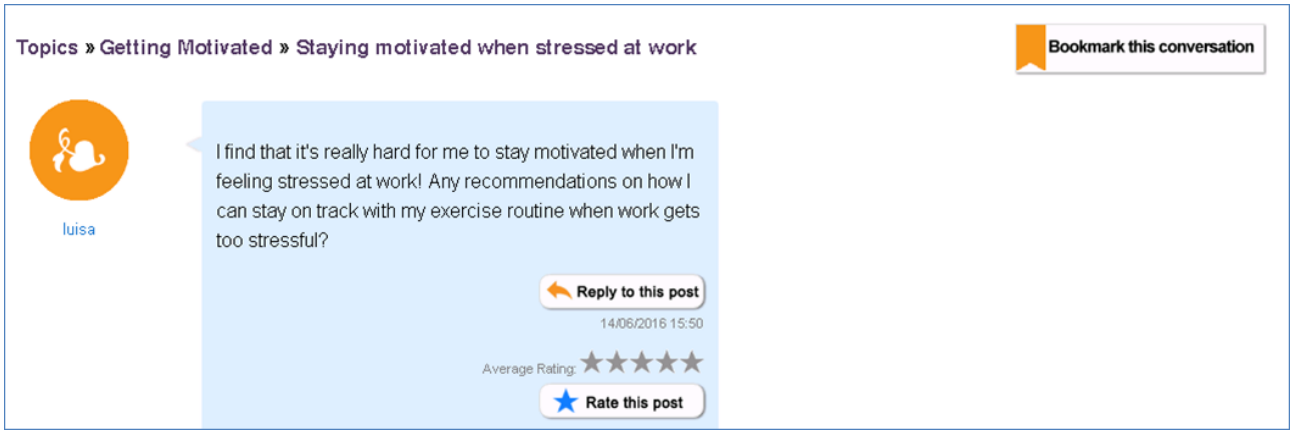
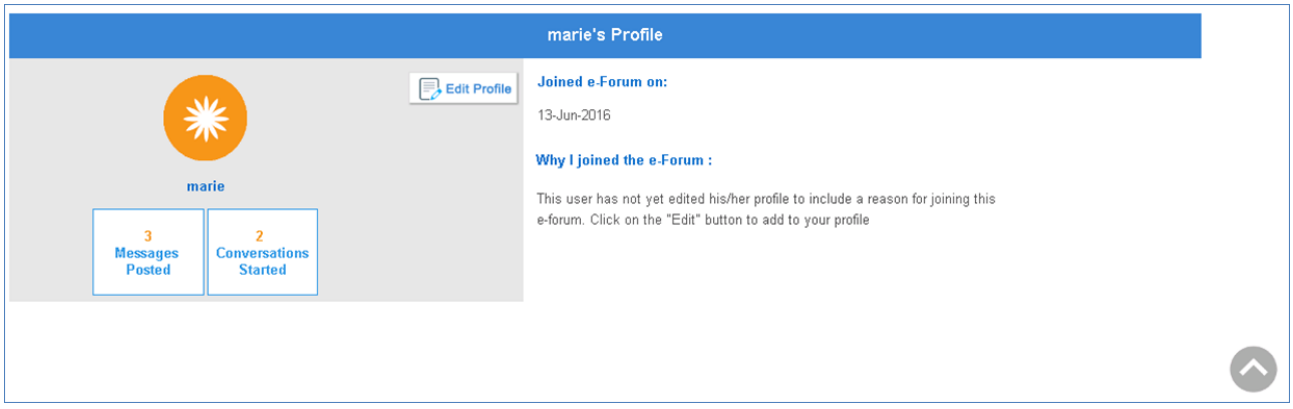


Figure 3. Final version of My Profile page. The final version of the "My Profile" page reflected user feedback in that it guided users to edit their profile by including both an icon and a verbal description of this button's function.



Objective

The objective of this study was to assess the usability of the user interface for our newly designed e-Forum. To achieve this goal, we obtained feedback from our target users (eg, cardiac patients) on key features including general feedback, overall user satisfaction, layout, navigation, and content of the e-Forum.

Methods

Study Design

An iterative design [26,28] examined the usability of the e-Forum, such that multiple groups of participants were asked to navigate the e-Forum. On the basis of analysis of feedback from each round, adjustments were made to the e-Forum; these changes were then assessed for usability with the succeeding group of new participants.

Participant Recruitment

Because the e-Forum aims to foster heart-healthy lifestyle changes that are applicable to all cardiac patients, including HF patients, we wanted to ensure that it was user-friendly to the wider, heterogeneous cardiac population. Thus, we recruited subjects from the Heart Function Clinic at the Peter Munk Cardiac Center, University Health Network. Patients were eligible to participate in this study if they were (1) male or female patients aged ≥18 years, (2) diagnosed with a CVD, including systolic HF with New York Heart Association Class I-III symptoms, and (3) fluent in English. To assess whether our e-Forum was easy to use for individuals with varying degrees of experience, we purposefully sampled an array of self-reported novice and advanced users of both computers and the internet. Individuals who did not use computers and the internet at all and were not willing to try these technologies were ineligible to participate in this study.

Procedure

This study received approval from the Research Ethics Board at the University Health Network. During the study visit, each consented participant was asked to complete a set of goal-oriented tasks and a feedback interview on the e-Forum. All study visits were completed within 1.5 hours.

Goal-oriented tasks were the same across all study rounds and included logging onto the website, watching a tutorial video, and using different features of the e-Forum (eg, editing sample user profiles, submitting, bookmarking, and rating sample posts; Multimedia Appendix 1). Instructions for each goal-oriented task were read to participants, before asking them to "think-aloud" as they completed each task [29]. This commonly used protocol allowed us to assess the ongoing thought processes and difficulties experienced by the users while using the program [29]. To prevent disruption in the think-aloud protocol, no guidance or assistance was provided during task completion, unless requested by the participant [29]. All participants were able to successfully complete the think-aloud protocol.

After completing the set of goal-oriented tasks, a semistructured interview was used to ask participants about their overall experience with the e-Forum and to allow them to make suggestions for its improvement in layout (eg, font size, colors, and formatting), navigation (eg, ease of use), and content (eg, highlighted topics, and features/functionalities, including bookmarking and rating functions). All think-aloud sessions and feedback interviews were audio-taped using a digital audio recorder and then transcribed verbatim for analysis. Finally, all subjects completed a demographics form and a user satisfaction questionnaire. The items on the user satisfaction questionnaire were based on the usability characteristics, as described by Nielson [30], and included a 5-point Likert scale (1="disagree very much"; 5="agree very much") asking participants to rate their level of satisfaction with different aspects of the e-Forum.

Data Analysis

After each study visit, a research assistant transcribed the audiotape verbatim, and a second research assistant independently compared this transcription with the audiotape to verify its accuracy. A content analysis of the transcripts from the study sessions identified themes related to the overall satisfaction and the layout, navigation, and content of the e-Forum. QSR NVivo (QSR International, Victoria, Australia) was used to manage the transcript data. Concurrent data collection and analysis and constant comparison [31] facilitated probing for further insights to confirm themes that arose in subsequent interviews [32]. Transcripts were independently coded by RT and AB, and divergent codes were discussed and resolved. Once the coding process was complete, a frequency count tallied participants' experiences in each theme [32]. Both quantitative frequency counts and qualitative interview excerpts were reported. Means, SDs, and percentages were calculated for data collected from the demographics and the satisfaction questionnaire forms.

Results

Participants

A total of 9 men and 4 women participated in this study over 3 rounds of data collection ($n_{round\ 1}$=5, $n_{round\ 2}$=5, and $n_{round\ 3}$=3). Saturation of the narrative data was obtained with this sample. Participants' age ranged from 32 to 81 years (mean=63.1, SD=13.8). The majority of participants were white (77%, 10/13), married (62%, 8/13), and had at least some postsecondary education (77%, 10/13). Six (46%, 6/13) participants were employed at the time of the study session. With regard to diagnosis, 7 participants (54%, 7/13) had been diagnosed with HF or cardiomyopathy, and 6 (46%, 6/13) had been diagnosed with other CVD, including cardiac amyloidosis, valvular heart disease, or ischemic heart disease (Table 1).

All 13 participants used the internet at home, with 12 accessing the internet via a computer and 1 via a mobile device. All participants reported at least being somewhat comfortable with computers and the internet. Nevertheless, there was variability in the degree to which participants used the computer/internet at home. Of the participants who had a computer at home, 50% (6/13) spent less than 5 hours per week on the computer, whereas the other half (6/13) spent more than 5 hours per week on the computer.

Table 1. Demographic characteristics of study participants.

Demographic variables	Round 1 (n=5), n (%)	Round 2 (n=5), n (%)	Round 3 (n=3), n (%)
Age in years			
30-49	1 (20)	2 (40)	0 (0)
50-69	3 (60)	1 (20)	2 (67)
>70	1 (20)	2 (40)	1 (33)
Gender			
Male	5 (100)	2 (40)	2 (67)
Female	0 (0)	3 (60)	1 (33)
Marital status			
Married/common-law	3 (60)	4 (80)	1 (33)
Single/separated/divorced	2 (40)	1 (20)	2 (67)
Highest education level			
High school	2 (40)	1 (20)	0 (0)
Some college/college	1 (20	1 (20)	1 (33)
Graduate/professional degree	2 (40)	3 (60)	2 (67)
Employment status			
Full-time/part-time	2 (40)	2 (40)	2 (67)
Retired/disability/leave of absence	3 (60)	3 (60)	1 (33)
Ethnicity			
White	3 (60)	4 (80)	3 (100)
Other	2 (40)	1 (20)	0 (0)
Diagnosis			
Heart failure/cardiomyopathy	1 (20)	4 (80)	2 (67)
Other cardiovascular disease	4 (80)	1 (20)	1 (33)

Similarly, although all participants had access to the internet at home, 54% (7/13) spent less than 5 hours per week on the internet, whereas 46% (6/13) spent more than 5 hours per week on the internet at home. Nevertheless, the majority also reported at least being somewhat comfortable with using online forums or message boards (62%, 8/13); and 5 participants (38%, 5/13) regularly used online forums or message boards for personal use (Table 2).

Overall Satisfaction and General Comments

Satisfaction With the e-Forum

Evaluation of the user satisfaction assessment indicated that, on average, participants in all 3 rounds were satisfied with their experience in using the e-Forum (Table 3). Similarly, the majority of participants made at least one comment regarding their overall satisfaction with the e-Forum, and the number of satisfactory comments per participant increased from 3.5 in round 1 to 5 in round 3. Unique comments included general statements of satisfaction, expressions of satisfaction with the moderated aspects of the forum, as well as satisfaction with the opportunity to connect with other patients with similar conditions (Tables 4 and 5).

Description of Use

Participants from all 3 rounds made a total of 41 individual comments describing how they would use the e-Forum. Participants indicated that they would use the e-Forum to exchange advice regarding lifestyle behavior change and to share/gather information regarding the management of their cardiac condition. They also indicated that they might enlist the help of family members when using the e-Forum, and that this interface may also be used to provide additional support for family members of cardiac patients. Participants said that other cardiac patients would also likely be interested in using the e-Forum for additional support and resources (Tables 4 and 5).

Potential Barriers

Participants in all rounds also speculated that there might be potential barriers to accessing or using the e-Forum for other cardiac patients. There was a decrease in the total number of comments made regarding potential barriers between round 1 (12 comments) and round 3 (4 comments). Potential barriers included lack of access to the internet, poor computer skills, and self-consciousness about typing or general ability to use computers. Other barriers included the potential unwillingness of some cardiac patients to share their experiences in managing their condition (Tables 4 and 5).

Table 2. Self-reported computer and internet use.

Computer and internet usage variables	Round 1 (n=5), n (%)	Round 2 (n=5), n (%)	Round 3 (n=3), n (%)
Use of computer: work			
Yes	3 (60)	2 (40)	2 (67)
No	0 (0)	0 (0)	1 (33)
Not applicable	2 (40)	3 (60)	0 (0)
Use of internet: work			
Yes	3 (60)	2 (40)	3 (100)
Not applicable	2 (40)	3 (60)	0 (0)
Use of computer: home			
Yes	5 (100)	5 (100)	2 (67)
No	0 (0)	0 (0)	1 (33)
Hours spent on computer: home			
<5 hours per week	4 (80)	1 (20)	1 (33)
>5 hours per week	1 (20)	4 (80)	1 (33)
Not applicable	0 (0)	0 (0)	1 (33)
Use of internet: home			
Yes	5 (100)	5 (100)	3 (100)
Hours spent on internet: home			
<5 hours per week	3 (60)	2 (40)	2 (67)
> 5 hours per week	2 (40)	3 (60)	1 (33)
Use of online forums/message boards for personal use			
Yes	1 (20)	2 (40)	2 (67)
No	4 (80)	3 (60)	1 (33)
Level of comfort: computers			
Somewhat or comfortable	4 (80)	3 (60)	1 (33)
Very comfortable	1 (20)	2 (40)	2 (67)
Level of comfort: internet			
Comfortable	4 (80)	3 (60)	1 (33)
Very comfortable	1 (20)	2 (40)	2 (67)
Level of comfort: online forums/message boards			
Not at all comfortable	1 (20)	0 (0)	0 (0)
Somewhat or comfortable	3 (60)	3 (60)	1 (33)
Very comfortable	0 (0)	0 (0)	1 (33)
Not sure	1 (20)	2 (40)	1 (33)

Navigation: Task Navigation and Comments

Task Navigation

All participants were able to successfully navigate the e-Forum, with correct navigations per participant increasing from 13.8 in round 1 to 14.7 in round 3. Successful navigation included the ability to complete the specific steps to use the various features of the forum (eg, logging on, playing the tutorial video, editing profiles, and submitting and managing posts). Each participant also made at least one navigation error during the course of the study session. Nevertheless, the average number of navigation errors per participant decreased across the 3 rounds (5 in round 1 to 3.7 in round 3). Common navigation errors included difficulty finding the “edit profile,” “rate this post,” and “bookmark” buttons because of button placement or poor labeling (Table 4). Common navigation errors were addressed in changes made to the e-Forum between each round. See below for details.

Table 3. User satisfaction with e-Forum.

User satisfaction assessment	Round 1 (n=5), mean (SD)	Round 2 (n=5), mean (SD)	Round 3 (n=3), mean (SD)
I learned how to use the forum quickly and easily	4 (0.6)	5 (0.6)	5 (0.6)
I can find the information I am looking for on the forum with no problems	4 (1.3)	5 (0.6)	5 (0.6)
I can make and reply to posts on the forum with no problems	4 (0.8)	4 (0.9)	5 (0.6)
I am confident that I can remember how to get around the forum on my own every time I log on	4 (0.6)	4 (0.6)	4 (1.0)
If I get lost on the forum, I am confident that I can find my way again	4 (0.6)	4 (0.8)	4 (1.0)
I am satisfied with the forum	4 (1.1)	4 (0.0)	4 (0.6)
I would use the forum regularly to help me better manage my heart condition	5 (0.5)	4 (1.1)	4 (0.6)

Table 4. Content analysis of usability of e-Forum.

Themes	Round 1 (n=5)				Round 2 (n=5)				Round 3 (n=3)			
	C/I[a], n	Ucs[b], n	P[c], n	Mean # of C/I per P[d]	C/I, n	Ucs, n	P, n	Mean # of C/I per P	C/I, n	Ucs, n	P, n	Mean # of C/I per P
Content												
Positive comments	14	5	4	3.5	10	3	5	2.0	8	2	3	2.7
Negative comments	8	5	3	2.7	16	9	4	4.0	4	3	2	2.0
Neutral comments	19	12	5	3.8	20	12	4	5.0	14	5	3	4.7
Navigation												
Positive comments	12	2	4	3.0	9	2	5	1.8	5	2	2	2.5
Negative comments	1	1	1	1.0	4	2	3	1.3	1	1	1	1.0
Layout												
Positive comments	15	5	5	3.0	21	5	5	4.2	10	4	3	3.3
Negative comments	10	6	2	5.0	15	8	4	3.8	1	1	1	1.0
Task navigation												
Correct navigation	69	17	5	13.8	70	16	5	14.0	44	16	3	14.7
Navigation errors	21	13	5	4.2	19	9	5	3.8	11	7	3	3.7
General feedback												
Satisfaction with forum	14	3	4	3.5	9	3	5	1.8	15	3	3	5.0
Potential barriers	12	4	4	3.0	5	2	3	1.7	4	1	2	2.0
Description of use	18	5	5	3.6	15	4	5	3.0	8	3	3	2.7

[a]Comments or incidents.

[b]Unique comments.

[c]Participants reported.

[d]Comments or incidents per participant.

Positive Navigation Comments

The majority of participants (85%,11/13) gave positive feedback (26 total positive comments) with regard to their ability to navigate the e-Forum. Positive comments included expressions of overall ease of navigation and indications that participants found the e-Forum easier to navigate or to understand as they used it (Tables 4 and 5).

Negative Navigation Comments

At least one participant in all 3 rounds provided a minimum of one negative comment on their ability to navigate the e-Forum. A total of 6 negative navigation comments were made, including overall difficulty with navigation and indications that the e-Forum was too complex to navigate (Tables 4 and 5). Nevertheless, all participants were able to successfully complete study tasks with little or no assistance.

Table 5. Sample comments of each of the themes.

Themes	Examples
Content	
Positive comments	"...the [highlighted] topics...active living, eating healthy, get motivated,...I feel these are all the topics that...people would be interested in." [B5, round 2]
	"...I'm...assuming that with every reply, I'll get something in my inbox as well, so that's...good...[because]...after its been forwarded it'll just send me an email and then I'll know okay, someone's replied and...have the answer to my question okay...that's good." [B4, round 2]
Negative comments	"I would definitely use [the forum] if there was more...information and the content was more rich." [A5, round 1]
	"...it wasn't clear to me what criteria I was supposed to use [to rate posts]...[there] was a bit of guess work involved in there." [A2, round 1]
Neutral comments	"...I would make it mobile friendly, I don't know if it's a mobile friendly site." [B5, round 2]
	"Another [topic] is sleeping...Sleeping is critically important...I would find it really interesting to understand how people approach that...what they think are good rules to follow, how they're doing it..." [C1, round 3]
Navigation	
Positive comments	"It's very clear...I think once you go through it once or twice, it [is] very simple to...follow." [A4, round 1]
	"This is very user friendly; I don't think navigating it is a problem...It's fairly intuitive and easy to follow. I don't think anyone who uses the internet regularly should have any difficulty...navigating it." [A5, round 1]
Negative comments	"It was a little difficult [to navigate]...I should say not very difficult because I've had an idea with the keyboard and I've looked at [something] similar to this...but [for] some people it may be very intimidating for them." [B3, round 2]
	"...there's too many pop up boxes...there's too many steps...[for something that] could be 1 step there's...3 steps instead...and I feel like people will get confused..." [B5, round 2]
Layout	
Positive comments	"I think [the layout is] very simple and nice; I like the simplicity, I like the use friendliness." [A5, round 1]
	"It's really well done, like the font and the color, and when you need to know something, it pops up where it needs to be...all of the buttons are great, when you touch them they work the first time...I think it's an excellent website." [C2, round 3]
Negative comments	"[I liked least] the cumbersome aspects of the webpage, having to click on that edit button which I didn't know was an edit button in the first place, unless I roll my mouse over it..." [B1, round 2]
	"...[the layout is] a little too busy; keep it simple is the right idea." [A1, round 1]
General feedback	
Satisfaction with forum	"The internet is a source of a lot of useful information but can also be a source of a lot of...misinformation, so...the moderation must be there, otherwise you [end up] working against your own best interests." [A2, round 1]
	"I think [this forum]...would come in handy [for other heart patients] ... [to] check on how they're doing, and how other people are doing. And it's pretty easy to use it on a computer." [C3, round 3]
Potential barriers	"When it comes to personal life...like health... [people] are not as open, for whatever reason...sometimes they don't want to talk about it, they just want to leave it alone." [A3, round 1]
	"There [are] some people that…wouldn't use it at all, just [because] they don't have a computer [or] maybe they're not going to adapt to a computer program and find it very difficult [and] intimidating." [A4, round 1]
Description of use	"Yeah [I would use a forum like this]. I would go in and see once you start having situations...with your health you see what other people are doing for exercising, eating healthy…" [A3, round 1]
	"It's the time in between your [appointments] when you have all the questions [about your diagnosis or procedure]...so having something as a resource to refer to would be something good." [B4, round 1]

Content

Positive Content Comments

A majority of participants (92%, 12/13) provided positive feedback regarding the content presented in the e-Forum. Participants indicated that they were satisfied with features/functionality of the e-Forum (eg, appreciation of confirmation messages after submissions, spell-checking, bookmarking, or tool tips) as well as the sample information provided (eg, indication that video and highlighted topics were helpful; Tables 4 and 5).

Negative Content Comments

Negative comments regarding the content of the e-Forum were made in each round, with a total of 9 participants making 28 such comments throughout the course of the study. Negative content feedback included dissatisfaction with certain features

or functionalities (eg, unclear rating criteria, tutorial video being overwhelming, and lack of spell-checking feature) and with the sample information provided in the e-Forum (eg, finding certain highlighted topics not relevant to their experience or that content was not comprehensive enough; Tables 4 and 5).

Neutral Content Comments

All participants made at least one neutral comment about the content of the e-Forum. Neutral comments included suggestions for additional features or functionalities of the e-Forum (eg, suggestions to create a search button or to host live support groups), suggestions for information to be provided on the e-Forum (eg, suggestions for additional highlighted topics or videos), and suggestions to create different forum groups based on varying health status (eg, diagnoses or lifestyles; Tables 4 and 5).

Layout

Positive Layout Comments

Every participant made at least one positive comment regarding the layout of the e-Forum. In total, 46 positive layout comments were made. Positive comments included satisfaction with buttons (eg, appropriate size and color), with font size (eg, easy to read or see), with colors (eg, attractive), and with the overall layout of the forum (eg, simple and easy to use; Tables 4 and 5).

Negative Layout Comments

At least one participant from each round expressed a negative comment regarding the layout of the e-Forum. However, such comments decreased from 5 to 1 comment per participant from round 1 to round 3. Negative comments included expressions of dissatisfaction with overall layout of the e-Forum (eg, layout too complex), font size (eg, too small or inconsistent), colors (eg, inconsistent or dated), or buttons (eg, not clearly labeled; Tables 4 and 5).

Changes Made to Forum

From Round 1 to Round 2

On the basis of the feedback provided by participants in round 1, various changes were made to the design of the e-Forum to improve usability for the subsequent round of participants. For example, buttons were moved and/or renamed to enhance visibility and accessibility. Text boxes were reformatted, and tool tips and button labels were changed or added (eg, changing “Return to Forum” to read “Go Back”) throughout the forum to improve the accessibility of associated features or functions. Finally, suggested changes were made to the layout of the forum, including changes in background and font color.

From Round 2 to Round 3

On the basis of the feedback provided by participants in round 2, more changes were made to the e-Forum to improve usability. Buttons were moved and/or renamed to enhance visibility and accessibility, and they were reformatted to improve consistency in layout. A keyword search feature was added to the e-Forum. Grammar and spell-checking features were added to textboxes; contact information for the research team, including expected response times, was also added to the e-Forum. See Figures 1-3 for a sample of the final version of the e-Forum at the completion of this usability study.

Discussion

Principal Findings

The e-Forum was designed to facilitate the establishment of a reliable and accessible online community for cardiac patients. This usability study was conducted to ensure that our e-Forum was user-friendly and accessible to our target patient population. An iterative design was used such that after each round of study sessions, changes were made to the e-Forum in response to participant feedback. Feedback included general reflections of user experiences as well as positive, negative, and neutral comments on the content, navigation, and layout of the e-Forum.

Overall, participants across all 3 rounds were highly satisfied with the e-Forum. Between rounds 1 and 3, expressions of satisfaction with the e-Forum increased, and fewer potential barriers were reported. Participants indicated that it would be helpful to speak with other cardiac patients and that they were particularly satisfied with the moderated aspect of the e-Forum. Participants indicated that they would use the e-Forum to exchange lifestyle behavior advice and general information regarding their health management with other patients. Having the moderation feature reassured them that the information they obtained would be reliable and safe. They also predicted that their family members would likely use the e-Forum on their own or together with the patients to obtain information and support.

As improvements were made to the e-Forum based on participant feedback, positive comments related to layout increased from the initial to the final round, whereas negative layout comments decreased. Moreover, negative comments related to content and the number of navigation errors decreased between rounds 1 and 3. These outcomes indicated that modifications made to the layout (eg, changes in colors and font sizes), as well as the content (eg, changes in descriptors and features, including the addition of the keyword search) of the e-Forum, likely improved the overall user experience and ease of use when interacting with our online community.

Limitations

The results of this study indicate that our e-Forum would likely be accessible to a diverse array of cardiac patients. However, there are some limitations to consider as efforts are made to disseminate this e-Forum to the wider patient population. For example, although the age of participants in this study was well representative of the target user population (10 participants were aged older than 50 years), those who agreed to participate in this study were primarily white, with at least some postsecondary education, and with self-reported experience and comfort using computers and the internet.

It is possible that our findings may be limited in generalizability, as the overall population of cardiac patients is more culturally and educationally diverse. Nevertheless, feedback provided by participants in this study also suggested that individuals from diverse backgrounds may actually be *more* comfortable asking questions about lifestyle behaviors on our e-Forum. These

comments are congruent with other studies that have found that users from rare or geographically dispersed backgrounds may be more likely to feel confident in exchanging experiences and advice with regard to their health management within an online community [33]. Similarly, although some participants suggested that the wider patient population may be less comfortable with or have limited access to the internet, such concerns may not be relevant, as it has been established that the majority of individuals in North America have access to the internet [34,35].

Future Directions

Given the limitations of this study, future studies may work to recruit a more diverse sample of patients to ensure ease of use of the e-Forum across a wide range of patient demographics and experiences. Future studies may also use back-end analytics to assess how participants organically use the e-Forum (eg, how often, for how long, and which features they use most frequently) to gather additional information about how best to maximize the usability of the e-Forum. From a design perspective, future versions of the e-Forum may also increase usability by programming additional features, including making the e-Forum more mobile-friendly, including of speech recognition software, creating additional tutorial videos, allowing users to change font sizes, and offering the e-Forum in multiple languages. Moreover, it will be important to assess the e-Forum's ability to ultimately promote continued user engagement in Web-based lifestyle counseling programs and long-term self-care adherence.

Conclusions

In this study, we found evidence to support the usability of our newly designed e-Forum. After each study round, changes were made to the e-Forum based on user feedback. For example, buttons were moved and/or renamed to enhance visibility and accessibility and features, including but not limited to, a keyword search, and tool tips were added throughout the e-Forum. As a result, a diverse sample of cardiac patients, in terms of age and self-reported comfort with computers/internet, were able to successfully navigate the e-Forum. Moreover, these users indicated satisfaction with the layout and content of the e-Forum and expressed interest in using this tool for practical and emotional support in managing their CVD. The high user satisfaction ratings indicate that the e-Forum provided an acceptable user experience. In sum, these findings support this tool and its potential role in promoting long-term lifestyle behavior change when paired with existing e-counseling programs, such as the CHF-CePPORT program [10].

Acknowledgments

The CHF-CePPORT trial is funded by an operating grant [#MOP-287717] from the Canadian Institutes of Health Research. The authors thank the participants for their contribution to this study.

Conflicts of Interest

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

Multimedia Appendix 1

Task instructions.

[PDF File (Adobe PDF File), 40KB - humanfactors_v5i2e20_app1.pdf]

References

1. Writing Group Members, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, American Heart Association Statistics Committee, Stroke Statistics Subcommittee. Executive Summary: Heart Disease and Stroke Statistics--2016 Update: a report from the American Heart Association. Circulation 2016 Jan 26;133(4):447-454. [doi: 10.1161/CIR.0000000000000366] [Medline: 26811276]
2. Evangelista LS, Shinnick MA. What do we know about adherence and self-care? J Cardiovasc Nurs 2008;23(3):250-257 [FREE Full text] [doi: 10.1097/01.JCN.0000317428.98844.4d] [Medline: 18437067]
3. Angus JE, Rukholm E, Michel I, Larocque S, Seto L, Lapum J, et al. Context and cardiovascular risk modification in two regions of Ontario, Canada: a photo elicitation study. Int J Environ Res Public Health 2009 Sep;6(9):2481-2499 [FREE Full text] [doi: 10.3390/ijerph6092481] [Medline: 19826558]
4. Grace SL, Barry-Bianchi S, Stewart DE, Rukholm E, Nolan RP. Physical activity behavior, motivational readiness and self-efficacy among Ontarians with cardiovascular disease and diabetes. J Behav Med 2007 Feb;30(1):21-29. [doi: 10.1007/s10865-006-9080-5] [Medline: 17109217]
5. Nolan RP, Upshur RE, Lynn H, Crichton T, Rukholm E, Stewart DE, et al. Therapeutic benefit of preventive telehealth counseling in the Community Outreach Heart Health and Risk Reduction Trial. Am J Cardiol 2011 Mar 1;107(5):690-696. [doi: 10.1016/j.amjcard.2010.10.050] [Medline: 21215382]
6. Durrani S, Irvine J, Nolan RP. Psychosocial determinants of health behaviour change in an e-counseling intervention for hypertension. Int J Hypertens 2012;2012:191789 [FREE Full text] [doi: 10.1155/2012/191789] [Medline: 22229083]

7. Nolan RP, Liu S, Shoemaker JK, Hachinski V, Lynn H, Mikulis DJ, et al. Therapeutic benefit of internet-based lifestyle counselling for hypertension. Can J Cardiol 2012 May;28(3):390-396. [doi: 10.1016/j.cjca.2012.02.012] [Medline: 22498181]
8. Walker RL, Gee ME, Bancej C, Nolan RP, Kaczorowski J, Joffres M, et al. Health behaviour advice from health professionals to Canadian adults with hypertension: results from a national survey. Can J Cardiol 2011;27(4):446-454. [doi: 10.1016/j.cjca.2011.03.013] [Medline: 21684718]
9. Nolan RP, Liu S, Feldman R, Dawes M, Barr S, Lynn H, et al. Reducing risk with e-based support for adherence to lifestyle change in hypertension (REACH): protocol for a multicentred randomised controlled trial. Br Med J Open 2013;3(8):e003547 [FREE Full text] [doi: 10.1136/bmjopen-2013-003547] [Medline: 23965936]
10. Nolan RP, Payne AY, Ross H, White M, D'Antono B, Chan S, et al. An internet-based counseling intervention with email reminders that promotes self-care in adults with chronic heart failure: randomized controlled trial protocol. JMIR Res Protoc 2014 Jan 30;3(1):e5 [FREE Full text] [doi: 10.2196/resprot.2957] [Medline: 24480783]
11. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
12. Jibaja-Weiss ML, Volk RJ. Utilizing computerized entertainment education in the development of decision aids for lower literate and naïve computer users. J Health Commun 2007 Oct;12(7):681-697. [doi: 10.1080/10810730701624356] [Medline: 17934944]
13. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
14. 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]
15. Trinacty M, Farrell B, Schindel TJ, Sunstrum L, Dolovich L, Kennie N, et al. Learning and networking: utilization of a primary care listserv by pharmacists. Can J Hosp Pharm 2014 Sep;67(5):343-352 [FREE Full text] [Medline: 25364016]
16. Seidel RL, Jalilvand A, Kunjummen J, Gilliland L, Duszak R. Radiologists and social media: do not forget about Facebook. J Am Coll Radiol 2018 Jan;15(1 Pt B):224-228. [doi: 10.1016/j.jacr.2017.09.013] [Medline: 29132999]
17. Bonniface L, Green L. Finding a new kind of knowledge on the HeartNET website. Health Info Libr J 2007 Dec;24 Suppl 1:67-76. [doi: 10.1111/j.1471-1842.2007.00742.x] [Medline: 18005296]
18. Lau AY, Dunn A, Mortimer N, Proudfoot J, Andrews A, Liaw S, et al. Consumers' online social network topologies and health behaviours. Stud Health Technol Inform 2013;192:77-81. [Medline: 23920519]
19. Gammon D, Strand M, Eng LS, Børøsund E, Varsi C, Ruland C. Shifting practices toward recovery-oriented care through an e-recovery portal in community mental health care: a mixed-methods exploratory study. J Med Internet Res 2017 May 02;19(5):e145 [FREE Full text] [doi: 10.2196/jmir.7524] [Medline: 28465277]
20. Gill PS, Whisnant B. A qualitative assessment of an online support community for ovarian cancer patients. Patient Relat Outcome Meas 2012;3:51-58 [FREE Full text] [doi: 10.2147/PROM.S36034] [Medline: 23185122]
21. Huh J, McDonald DW, Hartzler A, Pratt W. Patient moderator interaction in online health communities. AMIA Annu Symp Proc 2013;2013:627-636 [FREE Full text] [Medline: 24551364]
22. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [doi: 10.1016/j.amepre.2007.01.016] [Medline: 17466822]
23. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [doi: 10.1016/j.socscimed.2005.11.016] [Medline: 16361029]
24. Vissenberg C, Stronks K, Nijpels G, Uitewaal PJ, Middelkoop BJ, Kohinor MJ, et al. Impact of a social network-based intervention promoting diabetes self-management in socioeconomically deprived patients: a qualitative evaluation of the intervention strategies. Br Med J Open 2016 Apr 13;6(4):e010254 [FREE Full text] [doi: 10.1136/bmjopen-2015-010254] [Medline: 27075842]
25. Pai S, Boutin-Foster C, Mancuso CA, Loganathan R, Basir R, Kanna B. “Looking out for each other”: a qualitative study on the role of social network interactions in asthma management among adult Latino patients presenting to an emergency department. J Asthma 2014 Sep;51(7):714-719 [FREE Full text] [doi: 10.3109/02770903.2014.903967] [Medline: 24628522]
26. Wichansky AM. Usability testing in 2000 and beyond. Ergonomics 2000 Jul;43(7):998-1006. [doi: 10.1080/001401300409170] [Medline: 10929833]
27. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
28. Currie LM. Evaluation frameworks for nursing informatics. Int J Med Inform 2005 Dec;74(11-12):908-916. [doi: 10.1016/j.ijmedinf.2005.07.007] [Medline: 16099711]
29. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
30. Nielson J. Iterative user interface design. IEEE Computer 1993;26(11):32-41 [FREE Full text]

31. Patton MQ. Qualitative Research and Evaluation Methods, 3rd edition. Thousand Oaks, CA: Sage Publications Inc; 2002.
32. Neuendorf KA. The Content Analysis Guidebook. Thousand Oaks, CA: Sage Publications; 2002.
33. Reisner SL, Randazzo RK, White Hughto JM, Peitzmeier S, DuBois LZ, Pardee DJ, et al. Sensitive health topics with underserved patient populations: methodological considerations for online focus group discussions. Qual Health Res 2017 May 01:1049732317705355. [doi: 10.1177/1049732317705355] [Medline: 29298574]
34. Statistics Canada. 2014. Canadian Internet use and e-commerce data from the 2010 Canadian Internet use survey URL: http://publications.gc.ca/site/archivee-archived.html?url=http://publications.gc.ca/collections/collection_2016/statcan/11-627-m/11-627-m2014002-eng.pdf [accessed 2018-04-13] [WebCite Cache ID 6yeWeYbTk]
35. File T, Ryan C. 2014. Computer and Internet Use in the United States: 2013 (American Community Survey Reports) URL: https://www.census.gov/history/pdf/2013computeruse.pdf [accessed 2018-04-13] [WebCite Cache ID 6yeWvbmzN]

Abbreviations

C/I: comments or incidents
CHF-CePPORT: Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure
CVD: cardiovascular disease
HF: heart failure
P: participant
Ucs: unique comments

Edited by P Santana-Mancilla; submitted 28.08.17; peer-reviewed by L Gaytán-Lugo, S Fajardo, W Stut; comments to author 11.09.17; revised version received 07.02.18; accepted 27.02.18; published 18.05.18.

Please cite as:
Tanaka R, Banerjee A, Surikova J, Tracey J, Payne A, Ross H, Nolan R
A Moderated e-Forum for Adults With Cardiovascular Disease: Usability Study
JMIR Hum Factors 2018;5(2):e20
URL: http://humanfactors.jmir.org/2018/2/e20/
doi:10.2196/humanfactors.8820
PMID:29776901

Original Paper

Enhancing Home Health Mobile Phone App Usability Through General Smartphone Training: Usability and Learnability Case Study

Richard Harte[1,2,3], BEng, PhD; Tony Hall[4], BA, MIT, PhD; Liam Glynn[5], BA, MB, BCh, BAO; Alejandro Rodríguez-Molinero[1,6], MD, PhD; Thomas Scharf[7], PhD, FAcSS; Leo R Quinlan[2,3,8], BSc, PhD; Gearóid ÓLaighin[1,2,3], BE, MEngSc, PhD, CEng

[1]Electrical & Electronic Engineering, School of Engineering & Informatics, NUI Galway, University Road, Galway, Ireland

[2]Human Movement Laboratory, NUI Galway, University Road, Galway, Ireland

[3]CÚRAM, Centre for Research in Medical Devices, NUI Galway, University Road, Galway, Ireland

[4]School of Education, NUI Galway, University Road, Galway, Ireland

[5]General Practice, School of Medicine, NUI Galway, University Road, Galway, Ireland

[6]Clinical Research Unit, Consorci Sanitari del Garraf, Sant Pere de Ribes, Barcelona, Spain

[7]Irish Centre for Social Gerontology, Institute for Lifecourse and Society, NUI Galway, University Road, Galway, Ireland

[8]Physiology, School of Medicine, NUI Galway, Galway, Ireland

Corresponding Author:
Leo R Quinlan, BSc, PhD
Physiology
School of Medicine
NUI Galway
University Road
Galway,
Ireland
Phone: 353 91493710
Fax: 353 91494544
Email: leo.quinlan@nuigalway.ie

Abstract

Background: Each year, millions of older adults fall, with more than 1 out of 4 older people experiencing a fall annually, thereby causing a major social and economic impact. Falling once doubles one's chances of falling again, making fall prediction an important aspect of preventative strategies. In this study, 22 older adults aged between 65 and 85 years were trained in the use of a smartphone-based fall prediction system. The system is designed to continuously assess fall risk by measuring various gait and balance parameters using a smart insole and smartphone, and is also designed to detect falls. The use case of the fall prediction system in question required the users to interact with the smartphone via an app for device syncing, data uploads, and checking system status.

Objective: The objective of this study was to observe the effect that basic smartphone training could have on the user experience of a group that is not technically proficient with smartphones when using a new connected health system. It was expected that even short rudimentary training could have a large effect on user experience and therefore increase the chances of the group accepting the new technology.

Methods: All participants received training on how to use the system smartphone app; half of the participants (training group) also received extra training on how to use basic functions of the smartphone, such as making calls and sending text messages, whereas the other half did not receive this extra training (no extra training group). Comparison of training group and no extra training group was carried out using metrics such as satisfaction rating, time taken to complete tasks, cues required to complete tasks, and errors made during tasks.

Results: The training group fared better in the first 3 days of using the system. There were significant recorded differences in number of cues required and errors committed between the two groups. By the fourth and fifth day of use, both groups were performing at the same level when using the system.

Conclusions: Supplementary basic smartphone training may be critical in trials where a smartphone app–based system for health intervention purposes is being introduced to a population that is not proficient with technology. This training could prevent early technology rejection and increase the engagement of older participants and their overall user experience with the system.

(JMIR Hum Factors 2018;5(2):e18) doi:10.2196/humanfactors.7718

KEYWORDS

smartphone; aged; elderly; wearable electronic devices; telemedicine; user-computer interface; education; user centered-design; usability; connected health; human factors; falls detection

Introduction

Background

Digital mobile telephony potentially creates new opportunities to augment health care. Owing to their interactive features, large storage capacity, communication capabilities, and ability to access large knowledge databases, smartphones can present a novel means to deliver health care to individuals in the home. Consequently, smartphones are being used to deliver an increasingly wide range of personal health care solutions [1,2]. Although older adults have traditionally adopted new technology at lower rate than other age cohorts, the Pew Internet Research Center reports that the use of Internet technology by older adults is steadily increasing, with 2012 being the first year where more than half of people in the United States aged 65 years and older were using the Internet [3]. Recent studies have shown that older adults have a rich technology profile in terms of home appliances, TVs, PCs, and mobile phone apps and only differ from other technology using age groups in terms of Internet-based technology [4]. Although today's older adults have better uptake of mobile technologies than previous older adult groups [5], contemporary mobile devices such as smartphones still present a substantial challenge for older adults [6]. These challenges can be a result of numerous factors such as unfamiliarity or fear of technology, lack of perceived usefulness (PU), lack of perceived ease of use (PEoU), diminished interactive capabilities, and poor usability characteristics of the devices in question [7,8].

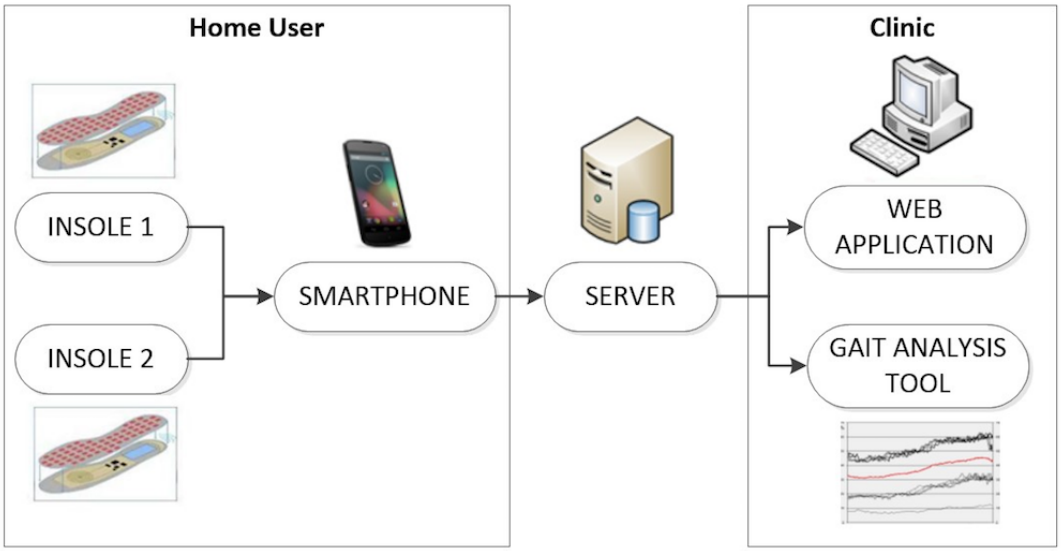
Methodological steps can be taken during the design process for these devices to ensure that the demands of the device do not exceed the capabilities of the older adult user (ref). For example, steps can be taken to ensure that interface elements such as buttons and text are usable, the device navigation can be designed to ensure that basic tasks only require a small number of steps, and the supporting documentation can be presented in an intuitive and simple manner [9-11]. However, these design aspects may not mitigate a new user's unfamiliarity with the device, and therefore, the potential for technology rejection may remain quite high [12]. This could have adverse effects when attempting to introduce smartphones to older adult users for the delivery of health care using an mHealth, telemedicine, or connected health infrastructure. In a previous study, a smartphone-based fall detection and prevention system was tested on a group of 39 older adult users over a 10-day period. Despite the system having undergone a full human-centered design process and the participants receiving adequate training on how to use the system, the system scored 70 of 100 on the System Usability Scale (SUS), indicating only average usability [13]. We suspected that unfamiliarity with smartphones and the specific demands of interaction with a smartphone, particularly the unique touch screen interactions required, may have led to poor usability outcomes. From this experience, we concluded that the outcomes of many trials and studies that involve the use of home health app design to run on smartphones could be compromised because of a lack of familiarity with the basic functioning of the device.

A period of pretrial introductory smartphone training in conjunction with concurrent recall-based learning tasks could present a potential solution to this problem. Effective training could provide a complete novice with a better chance of adopting the technology, thereby increasing the potential effectiveness of smartphone-based mHealth and connected health interventions for that person. In a study of a group of older adults who were being introduced to an mHealth pain management smartphone app, 61% of the participants cited "provide training on device use" as the main requirement if their potential use of the technology was to be enhanced and ultimately sustained, followed by 30% who cited "tailoring the device to the user's functional needs (ie, usability)" as the secondary requirement [14]. Therefore, we can conclude that twice as many participants felt that adequate introductory training was more important than enhancements to the design of the app in terms of usability. With this important finding in mind, this paper will provide an enumerative and detailed methodology to achieve this introductory training as efficiently as possible and therefore potentially mitigate any potential usability problems the technology may have.

Objectives

In this study, we trained 22 participants to use the same smartphone-based fall detection and prevention system, as described in the study by Stara et al [13], over a period of 5 days. This system is the Wireless Insole for Independent and Safe Elderly Living (WIISEL) system [15]. The system is designed to continuously assess fall risk by measuring various gait and balance parameters and is also designed to detect falls. The system is targeted at older adults who are at high risk of falling [16]. The system consists of a pair of instrumented insoles and a smartphone, which are worn by the user during daily activity. Data collected by embedded sensors in the insoles are sent to the smartphone and then uploaded via an Internet connection to a server in a clinic for processing and analysis. The data are presented in various ways to a specialist via a Web app and desktop-based gait analysis tool. The architecture of the system is illustrated in Figure 1.

Figure 1. The Wireless Insole for Independent and Safe Elderly Living (WIISEL) System.



The 22 participants were instructed to carry out a number of specific tasks with the WIISEL smartphone app (see the Methods section for details of these tasks). Half of the participants were provided with additional concurrent training in the general use of the smartphone, which began 2 days before the trial and continued. Our hypothesis was that the group that received the additional smartphone training would have a better user experience with the system.

Methods

Participants

Participants were recruited from the Galway city area. Twenty-two participants were recruited 74±5.5 years providing informed consent under ethical approval provided by University College Hospital Galway. Participants were split into 2 groups as outlined in Table 1. Whether a participant belongs to Group 1 (No extra training) or Group 2 (Extra training) was decided at random with 50% of participants belonging to each group.

All training was carried out in the participant's home by the lead researcher who followed the same protocol for each participant. The researcher visited the participants' home each day to teach them new tasks and to observe them carrying out previously learned tasks. A systematic cuing hierarchy approach and a "think aloud" protocol [17] were used by the lead researcher for the training procedure. These will be outlined in greater detail.

Technology Acceptance Indicators

To indicate how participants would fare with the introduction of the new technology, their current mobile technology capability was assessed. Participants were split into subgroups based on their previous experience with mobile technology to allow for further analysis. When it comes to classifying users based on their expertise or previous experience with mobile technology, there is no set classification system. Most so-called "expert" users are simply users who have gained hands-on experience of using the technology and may not have used a user manual for their device [18]. However, using an understanding of users' prior technical knowledge to predict their future adoption of technology is well supported [19]. We classified the participants into 3 separate categories based on their observed performance in carrying of a series of simple functional tasks with their mobile device. The functional tasks were chosen as representing real-world use requirements of the mobile device and were sufficiently challenging to highlight differences in skill level among user groups [18]. The 3 categories are outlined in Table 2.

In this context, "feature phone" is a retronym used to describe low-end mobile phones which are limited in capabilities in contrast to smartphones, that is, they are phones that have basic call and short message service functionality but do not have extensive media or Internet capabilities**.**

This procedure for practically dividing participants into technology experience groups is illustrated in the flowchart in Multimedia Appendix 1.

Perceived Ease of Use and Perceived Usefulness

PEoU and PU are the 2 key usability indicators pertaining to technology design. The influence of PEoU and PU on behavioral intent, and hence technology adoption, has been supported for the use of technology by older adults and specifically for the use of communication technology [20]. To measure whether there was any correlation between either PU or PEoU and the eventual usability outcomes, we measured the participant's PEoU and PU of smartphone technology before the trial started. We used a 7-point Likert scale using items from the technology acceptance model [21,22] to establish PU and PEoU.

Table 1. Participants were split into 2 groups based on what kind of training they would receive.

Group	Label	Training
1	No extra training	Only receive smartphone training which is necessary to operate the WIISEL[a] app. This group used the WIISEL system in their home for 5 days.
2	Extra training	This group also used the WIISEL system in the home for 5 days, but received extra smartphone training which began 2 days before starting the WIISEL trial and continued for the first three days of the WIISEL trial. The smartphone training was intended to make these participants familiar with the functions of the phone.

[a]WIISEL: Wireless Insole for Independent and Safe Elderly Living.

Table 2. Participants within each group were further classified based on their observed performance.

Category	Definition	Example
1 (novice user)	No experience with smartphones and basic capability with feature phone.	A user who does not own any sort of mobile phone OR a user who does own a feature phone but cannot demonstrate to satisfactory level that they can make a call, receive a call, and send or receive a text message without problems.
2 (intermediate or average tech user)	Perfect capability with feature phone, but limited capability with smartphones, may have been exposed to touch screen interfaces before.	A user who does own a feature phone and can demonstrate to a satisfactory level that they can make a call, receive a call, and send a text message OR a user who owns a smartphone and can demonstrate to a satisfactory level that they can make and receive a call but who cannot effectively send a text message.
3 (competent user or familiar with Internet technologies and mobile devices)	Adequate capability with (own) smartphone or with related touch screen devices (such as a tablet)	A user who owns a smartphone and can demonstrate satisfactorily that they can make or receive a call, receive a call, and send a text message OR a user who owns a tablet and can successfully send an email to the researcher.

Training Procedure

Our training procedure was based on an approach known as the errorless fading of cues technique [23]. This technique involves reducing the cues on repetitive tasks until the user can complete the task without error. The overall training schedule for Group 1 (No extra training) and Group 2 (Extra training) are outlined in Table 3. Training blocks are broken up to ensure that the participants were not overburdened with new training. Overall, each participant in Group 1 was subject to 0.75 to 1 hour of training or testing time per day, whereas each participant in Group 2 was subject to 1.5 to 2 hours of training or testing time per day. This included regular breaks and the time taken for the researcher to record metrics after each task. The number of days of training was chosen to be long enough to give participants the best chance of achieving some sort of mastery [23] but short enough to allow for convenience in having participants and trainer available for consecutive days.

The WIISEL specific tasks which were to be carried out by both Group 1 (No extra training) and Group 2 (Extra Training) are listed in Table 4. These tasks were selected based on a use case analyses of the WIISEL system [24] and were split into 2 different lesson blocks to reduce the burden on the participant by implementing small measurable objectives [25].

Table 3. Training schedule for each group. O indicate an introductory lesson, whereas X boxes indicate observational cue-assisted training. The dash (—) indicates no training on that day for that lesson block.

Group	Training type	Lesson block	Day −2	Day −1	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
1 (No extra training)	WIISEL[a] app training	Block 1	—	—	O	X	X	X	X	X
		Block 2	—	—	—	O	X	X	X	X
2 (Extra training)	Smartphone training	Block 1	O	X	X	X	X	X	—	—
		Block 2	O	X	X	X	X	X	—	—
		Block 3	—	O	X	X	X	X	—	—
	WIISEL app training	Block 1	—	—	O	X	X	X	X	X
		Block 2	—	—	—	O	X	X	X	X

[a]WIISEL: Wireless Insole for Independent and Safe Elderly Living.

Table 4. Wireless Insole for Independent and Safe Elderly Living (WIISEL) tasks.

Lesson block	Task number	Task title	Task description
1	1	Check System Status	Turn on phone from power-off state, enter WIISEL[a] app, and view system status
1	2	Connection Sequence	Enter app and carry out Connection Sequence to the WIISEL insoles
1	3	Upload Sequence	Upload WIISEL data from the app to the server
1	4	Minimize App	Minimize the app using the home button
1	5	Fall Detection Response	Respond to the fall alarm according to manual instructions
2	6	Reset Sequence	Carry out the app Reset Sequence via the smartphone settings
2	7	Log-in Sequence	Log-in to the app using the supplied username and password

[a]WIISEL: Wireless Insole for Independent and Safe Elderly Living.

Both groups were exposed to the WIISEL system over a course of 6 days (day 0 was training only, Days 1 to 5 were used to record performance, there was also some training on day 1, see step 2 in the following). This exposure consisted of the following steps:

1. On day 0 (first day of training), the researcher carried out a walkthrough for all the tasks in lesson block 1 (Table 4). For each task, the researcher walked the participant through the task and demonstrated it using the WIISEL user manual(s) as a reference (Multimedia Appendix 2). The participant repeated each task until no further cues were required. There was no recording of metrics on this day.
2. On day 1 (second day of training and first day of testing), the researcher carried out a walkthrough for all the tasks in lesson block 2 (Table 4), going through the same training routine as with lesson block 1 on day 0. The researcher also asked the participant to carry out the tasks in lesson block 1 by recalling their lessons from the previous day. The participant was instructed to try and complete the block 1 tasks without cues or input from the researcher, with the researcher providing cues only when it was clear that a cue was needed. The participant carried out each task 3 times.
3. At the end of each completed block 1 task, the user provided satisfaction ratings using the After Scenario Questionnaire (ASQ) [26]. Usability metrics such as task completion time, number of errors made, and number of cues required were recorded. The cuing hierarchy provided by the trainer comprised 5 different cue classifications [23], with 4=full explanation and demonstration; 3=the same verbal explanation as above but pointing to the next step before the participant executing it; 2=no verbal guidance provided, only pointing to the correct response before the participant executing it; 1=confirmation of a correct query, for example, “Do I tap details?”; and 0=no support provided. A lower score indicates greater proficiency with the device. Participants were given the standard instruction to think aloud [27].
4. From day 2 and day 5 on a daily basis, the participant carried out the WIISEL tasks from block 1 and block 2 (Table 4) under observation by the researcher. The researcher used the same metrics from day 1 (see step 3 above) to measure performance.
5. At the end of day 5, a semi-structured interview was carried out with each participant, and they also filled out the SUS questionnaire to provide an overall score of their user experience with the WIISEL system [28].

An example of the user manual for the WIISEL app is shown in Multimedia Appendix 2 and was the product of a comprehensive human-centered design process which tested and informed design changes of the WIISEL smartphone interface and the layout and content of the manual. Both Group 1 and Group 2 used this manual. Group 2 was also provided with smartphone-specific training in parallel with their WIISEL training. The smartphone-specific tasks are listed in Table 5. The smartphone tasks were selected based on recent studies on most popular usage patterns for smartphones [29,30].

This training was started 2 days before the WIISEL exposure began (we will refer to these days as day −2 and day −1). The routine was completed as follows:

1. On day −2, lesson blocks 1 and 2 (Table 5) were completed. For each task, the researcher walked the participant through the task and demonstrated it, using the NEXUS user manual(s) as a reference. The participant repeated each task until no further cues were required. There was no recording of metrics on this day.
2. On day −1, lesson block 3 (Table 5) was completed. Also on this day, the researcher asked the participant to try and recall their tasks from the previous day. The same recording of metrics was carried out as for the WIISEL training. This continued up to day 3.
3. It was seen as essential to gradually embed the task in its natural context with regular time constraints and less predictable occurrence [18]. Therefore, participants were asked to ring the lead researcher each day at a pre-agreed time and to send a short message service to accompany the in-situ observations [23].

The LG-Nexus 5 User Manual outlined in a step-by-step format how to complete basic tasks such as making phone calls and sending text messages (Multimedia Appendix 3). This manual was used for the Group 2 smartphone training.

Table 5. Smartphone tasks.

Lesson block	Task number	Task	Notes
1	1	Power and lock settings	Turn on phone from power-off state/unlock phone/lock phone
1	2	Dial a number to call	Dial a number into your phone and call it (present an arbitrary number)
1	3	Phone call	Receive a phone call and hang up and reject a phone call
2	4	Store a number in your phone and then call the stored number	Store a number in your phone, go to contacts to call the stored number
2	5	Text message	Read a text message/reply to the text message
3	6	Install an app	Install a television channel player (or similar) on the phone
3	7	Google search	Search for the term “Cinema Times Galway” in Google
3	8	Camera	Take a picture with the camera and go to gallery to see where the picture is stored

Analyses of Data

Data were interrogated using both qualitative and quantitative approaches. To compare how each group performed in terms of the WIISEL tasks (Table 4), *t* tests were used to seek statistical significance between groups for metrics such as errors made, cues required, completion time, and ASQ scores for each task. To reduce the effects of potential outliers for the task completion time metric (eg, a participant takes an unusually long time because of very slow typing or has to return to the beginning of the task because of a serious error), the logarithmic-based geometric mean was used [31]. Mean SUS scores were also compared for each group using *t* tests. Ethnographic observations were also made on the types of errors committed with the smartphone by reviewing notes and videos of the lessons.

Results

Overview

The results are presented in a series of stages. First, we will present our findings on the technology profiles of the participants. Next, we will compare the individual metrics of task times, errors made, cues required, and ASQ score between the two groups for each WIISEL task. Next, we will compare the SUS score for the two groups from their use of the WIISEL system. Finally, we will compare the results with some other usability studies that have been carried out with the WIISEL system.

Technology Profiles

The breakdown of participants into the different technology categories and age categories are presented in Table 6. The only significant correlation that was observed indicated that users with greater experience with mobile technology had a higher level of PU of smartphones. There was a minor correlation observed between increased technology experience and PEoU of smartphones. No significant correlation was found between age and PEoU (R=.2, P=.21) or between age and PU (R=−.04, P=.86). Weak positive correlation was found between technology experience and PEoU (R=.39, P=.15); Strong positive correlation was found between increased technology experience and PU (R=.6, P=.005).

Table 6 shows the results of the PE and PEoU questionnaire applied to each participant before the study began. The scale runs from 1 to 7, with 7 indicating a higher level of PU or PEoU. It provides a summary of correlation analysis between these questionnaires and age and technology categories of the participants.

In terms of mobile technology ownership, only 3 of the 22 participants (13%) did not own any sort of phone, whereas 6 (27%) owned a smartphone. The rest of the participants (60%) owned feature mobile phones, with the Nokia feature mobile phone (various models) proving the most popular. Five participants owned tablets although only 2 could be said to be proficient and frequent users. Of the 5 tablet users, none owned smartphones (all owned feature mobile phones). Therefore, the number of participants with some sort of touch screen experience was measured at 11 (50%).

Group Comparison of Wireless Insole for Independent and Safe Elderly Living Tasks

Task Completion Time

Although, in general, Group 1 took longer to complete tasks, particularly on days 1-3, none of these differences were found to be statistically significant (alpha=.05). It was observed that there was much wider variance in the task times for Group 1 than Group 2 (Figures 2 and 3), particularly for the early days of testing. By days 4 and 5, however, variances were relatively equal, and the difference in means was negligible.

Cues Required to Complete Task

The cues required for the routine tasks such as Connection Sequence and Upload Sequence were negligible for each group with no significant differences found (Figure 4).

Table 6. Number of participants who fell into each different age groups and technology experience categories (based on demonstration).

Variable	Number of participants (n=22)	PEoU[a] of smartphones	PU[b] of smartphones
Mean score	—	5.3	5
Age, in years			
65-69	5	5.3	4.7
70-74	7	5.2	5.1
75-79	5	4.7	5.6
80+	5	5	6
Technology category			
1 (novice)	6	4.4	4.5
2 (intermediate)	10	4.9	5.2
3 (expert)	6	5.6	6.3

[a]PEoU: perceived ease of use.

[b]PU: perceived usefulness.

Figure 2. No significant difference in mean task completion time between Group 1 and Group 2 was observed for the Connection Sequence (A) or the Upload Sequence (B). All times are shown in seconds.

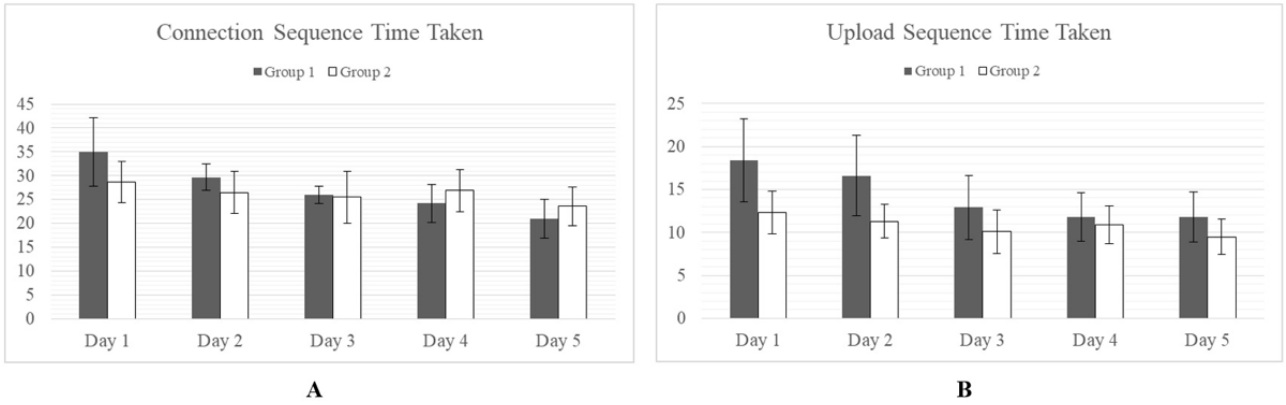


Figure 3. No significant difference in mean task completion time between Group 1 and Group 2 was observed for the Reset Sequence (A) or the Log-in Sequence (B). All times are shown in seconds.

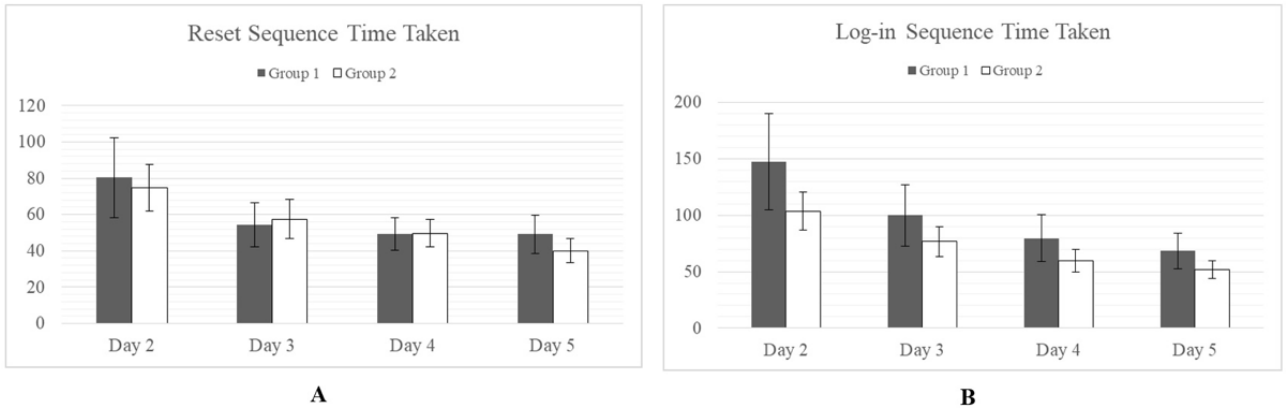
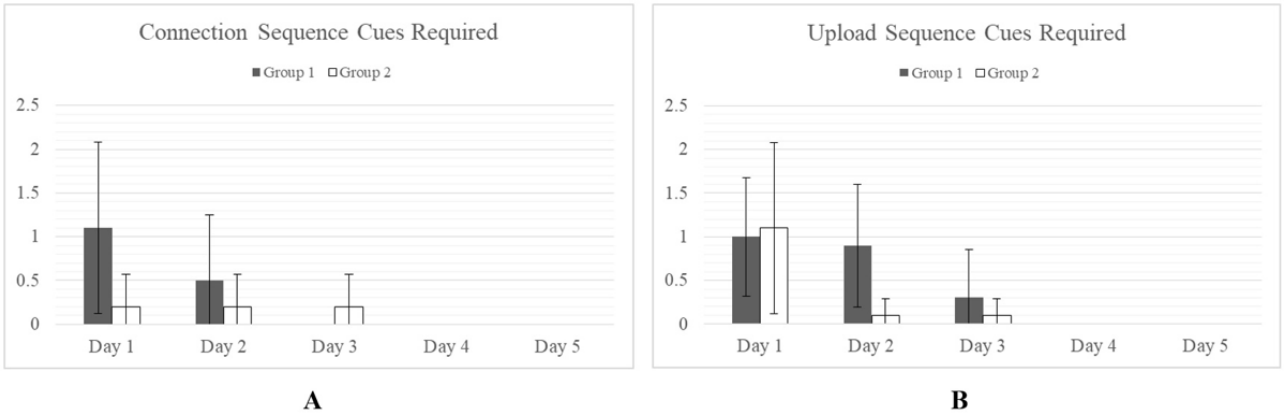


Figure 4. Cues required are measured as the total number of cue scores accumulated over the course of a task. No significant difference in cues required for Group 1 and Group 2 was observed for the Connection Sequence (A) or the Upload Sequence (B).



However, significant differences (alpha=.05) in cues required were observed for the more complex tasks, Reset and Log-in Sequences. Group 1 required more cues on average than Group 2 for the Reset and Log-in Sequences occurring on days 2 and 3 (Figure 5). By day 5, both groups had reached parity.

Errors Committed During Each Task

Errors were counted as when a user reached a point in a task where they could not continue without carrying out either a reversing action or required them to start the task again. No significant difference was observed between groups during the Connection Sequence and Upload Sequence (Figure 6).

Significant differences were observed. Statistically significant differences were observed between Group 1 and Group 2 for the Reset Sequence but not for the Log-in Sequence (Figure 7). By day 5, no difference was observed between groups for any task.

After Scenario Questionnaire Scores

ASQ scores show close agreement between groups on the ease of the Connection, Upload, and Reset Sequences (Figure 8).

The only statistically significant (alpha=.05) difference was seen in days 2 and 3 of the Log-in Sequence when Group 1 was shown to have scored significantly lower than Group 2 (Figure 9). By day 4, both groups had reached parity.

Overall User Experience

SUS scores showed significant differences (alpha=.05) between groups (Figure 10). In addition to measuring the total SUS, the scale was also split into its subscales to show learnability and usability scores. Group 1 versus Group 2 averages showed scores of 73 versus 87.25 (P=.007), 36.25 versus 63.75 (P=.04) and 82 versus 93 (P=.02) for SUS total, learnability, and usability, respectively.

Predictors of Positive User Experience

The regression analysis showed that there was a moderate to strong correlation between PEoU and PU and some of the SUS measures for each group. There was no significant correlation for age and technology experience as a predictor of SUS measures (Table 7).

The Impact of Performance Metrics on Overall User Experience

The regression analysis indicates that cues required, errors made, and the ASQ all influenced the SUS outcomes. Time taken was observed to be less of a factor, only showing strong correlation with learnability for Group 2 (Table 8).

Figure 5. Cues required are measured as the total number of cue scores accumulated over the course of a task. *P<.05. Statistically significant differences are observed between Group 1 and Group 2 for the more complex Reset Sequence (A) and Log-in Sequences (B). By day 5, however, no difference is observed between each group.

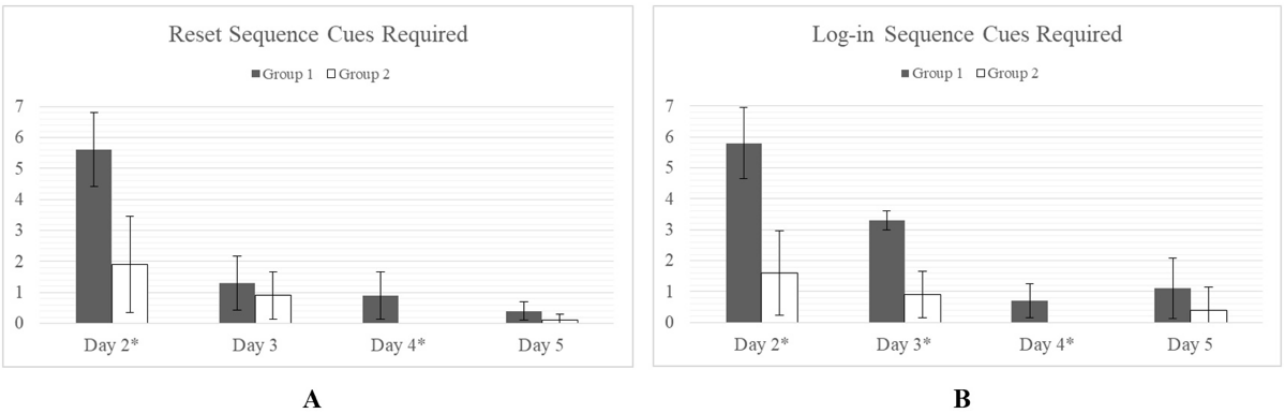


Figure 6. Average errors committed over the course of completing each task. No significant difference in errors committed between Group 1 and Group 2 was observed for the Connection Sequence (A) or the Upload Sequence (B).

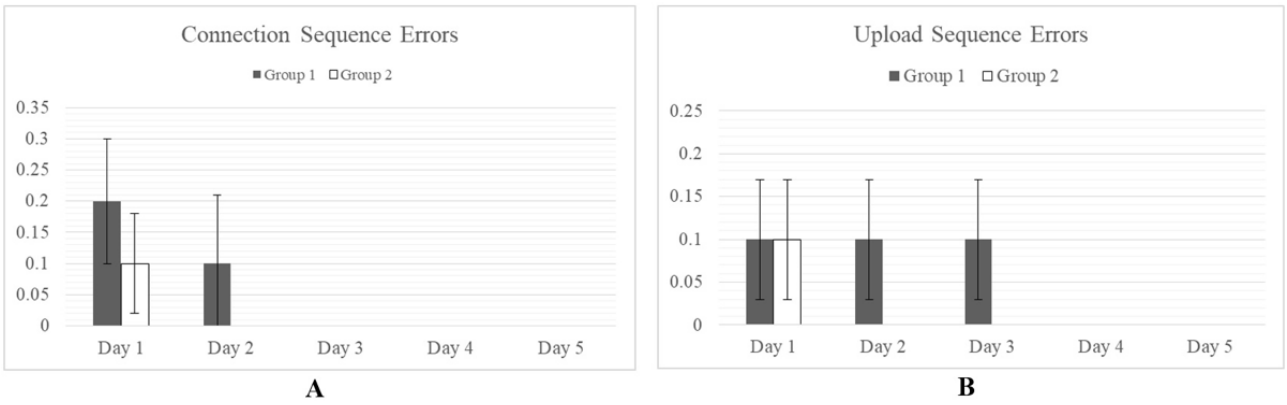


Figure 7. Average errors committed over the course of completing each task, *P<.05. Statistically significant differences are observed between Group 1 and Group 2 for the Reset Sequence (A). By day 5, no difference is observed between groups for any task.

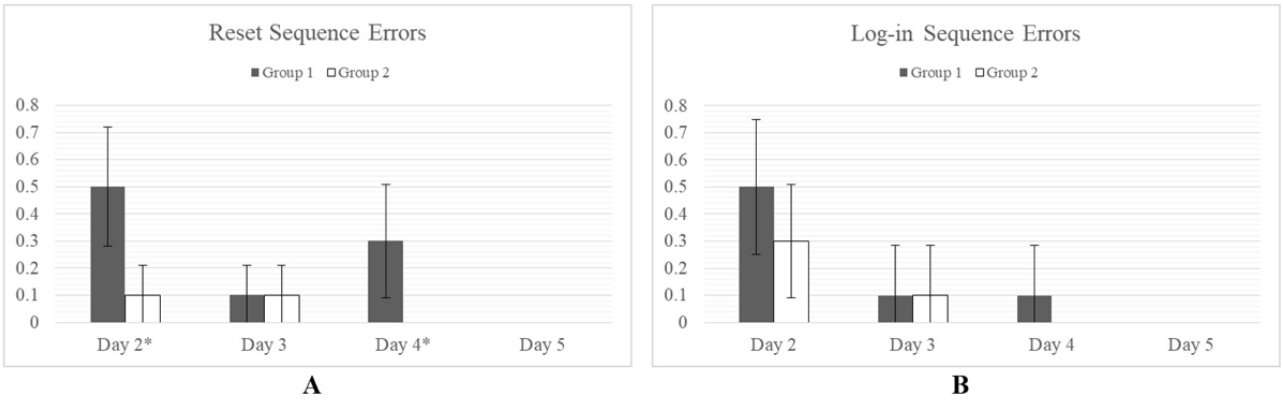
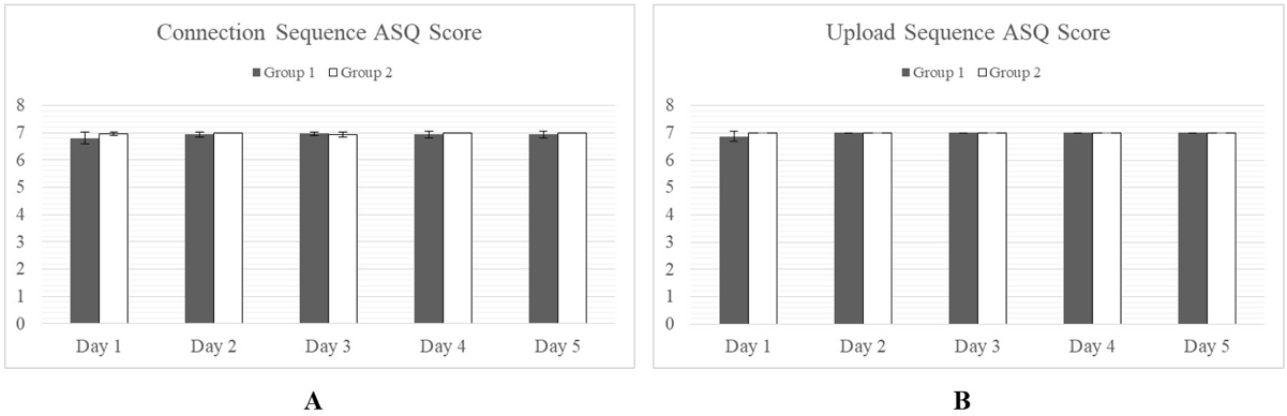


Figure 8. Satisfaction with the task in terms of ease, time taken, and supporting documentation. *P<.05. A score of 7 indicates maximum. No differences were observed for the Connection Sequence (A) or Upload Sequence (B) although significant differences were observed for the Log-in Sequence on days 2 and 3. ASQ: After Scenario Questionnaire.



Comparison of Results With Previous Wireless Insole for Independent and Safe Elderly Living Usability Studies

When comparing the SUS outcomes to previous studies of the WIISEL system, we can see how dramatic the effect of the supplementary smartphone training on the Group 2 participants in this study really was. In Figure 11, we can see that Group 1 exhibited similar SUS results to the participants who took part in a controlled usability test [24] and an open trial [32]. In both of these studies, the participants received no supplementary smartphone training. The Group 2 participants scored significantly higher than the other three groups.

Figure 9. Satisfaction with the task in terms of ease, time taken, and supporting documentation. *P<.05. A score of 7 indicates maximum. No differences were observed for the Connection, Upload, and Reset Sequences although significant differences were observed for the Log-in Sequence on days 2 and 3. ASQ: After Scenario Questionnaire.

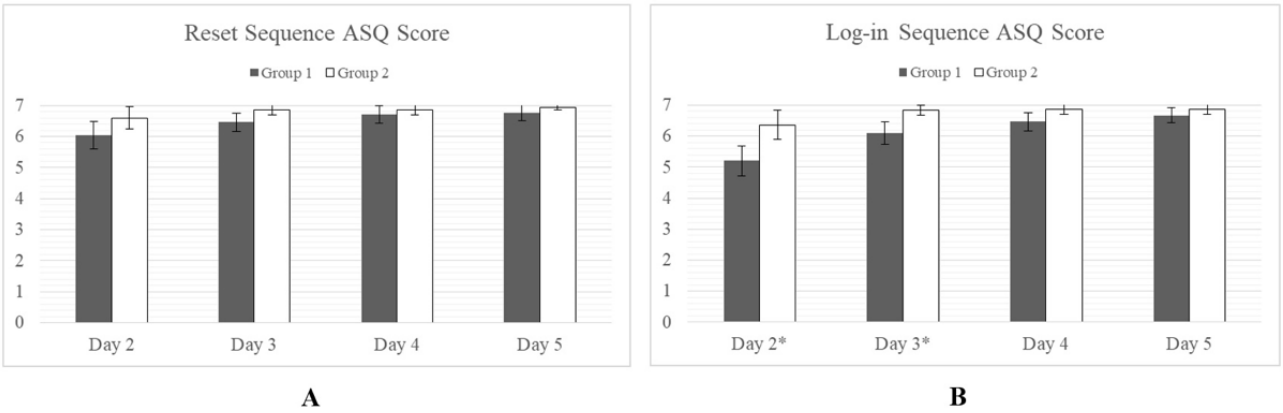


Figure 10. Satisfaction with the usability of the device. *P<.05, **P<.01. A score of 100 indicates maximum. Significant differences are observed between groups for each System Usability Scale (SUS) metric.

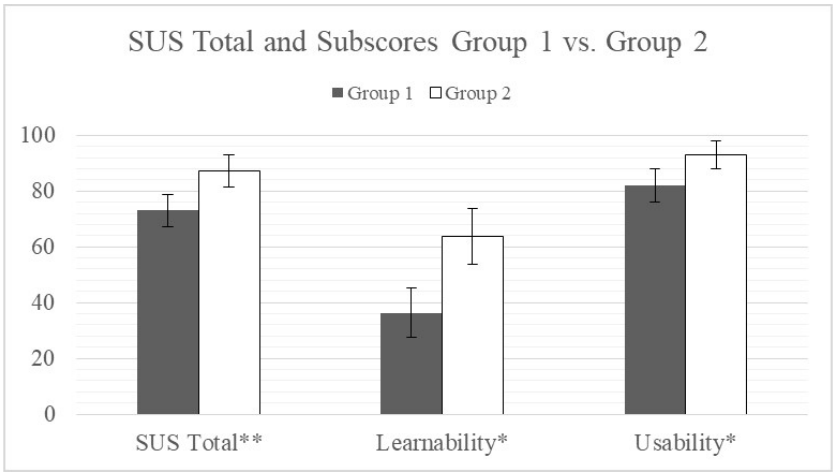


Table 7. Regression analysis of satisfaction measure against predictors such as technology experience, perceived usefulness, perceived ease of use, and age.

Group	System Usability Scale (SUS) measures	Technology category (experience)	Perceived usefulness	Perceived ease of use	Age
1 (No extra training)	Overall SUS	0.33	0.47	0.61[a]	0.12
	Usability subscale	0.27	0.19	0.6[b]	0.34
	Learnability subscale	0.33	0.66[b]	0.63[b]	0.51
2 (Extra training)	Overall SUS	0.35	0.57[a]	0.59[a]	0.19
	Usability subscale	0.33	0.69[b]	0.07	0.11
	Learnability subscale	0.18	0.133	0.82[c]	0.43

[a]*P*<.10.

[b]*P*<.05.

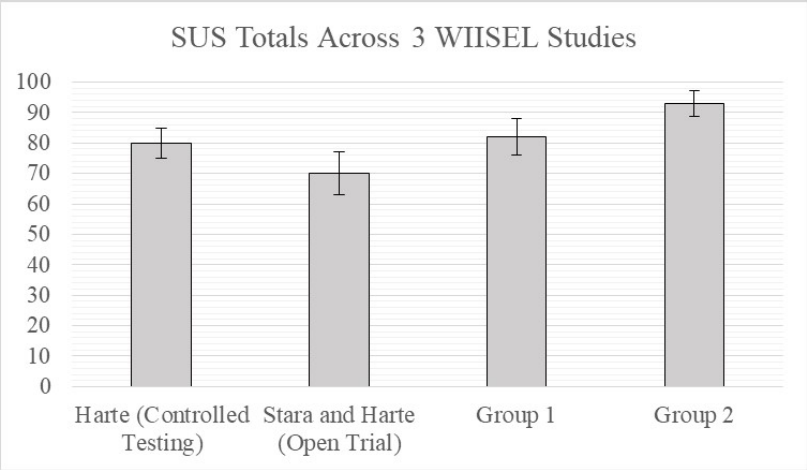
[c]*P*<.01.

Table 8. Regression analysis of satisfaction measure against performance metrics. SUS: System Usability Scale.

Group	SUS measures	Time taken	Cues required	Errors made	After Scenario Questionnaire
1 (No extra training)	Overall SUS	0.29	0.47	0.7[a]	0.634[b]
	Usability subscale	0.37	0.48	0.8[c]	0.63[b]
	Learnability subscale	0.07	0.78[c]	0.4	0.39
2 (Extra training)	Overall SUS	0.32	0.84[c]	0.44	0.42
	Usability subscale	0.1	0.7[b]	0.55[a]	0.21
	Learnability subscale	0.62[b]	0.54[a]	0.305	0.58[a]

[a]$P<.10$.
[b]$P<.05$.
[c]$P<.01$.

Figure 11. Comparison of System Usability Scale (SUS) scores from various assessments of the user experience of the WIISEL system. WIISEL: Wireless Insole for Independent and Safe Elderly Living.



Discussion

Review of Findings

This study aimed to understand how supplementary smartphone training could enhance an older adult's user experience of a smartphone-based connected health system (WIISEL). The results show that overall Group 2 (Extra Training) had a more positive user experience with the WIISEL system according to the statistically significant difference observed in the SUS (and its subscales) scores between the two groups at the end of the 5-day use period. We observed that in the first 3 days of WIISEL use, Group 2 outperformed Group 1 (No extra training) in key usability categories such as errors committed, cues required, and ASQ scores. No significant difference was observed in task times between each group. It was observed that by the fourth day of use, both groups were recording similar performance metrics, implying that there was a ceiling effect, above which no extra smartphone training could have any significant influence on WIISEL use performance. On the basis of the evidence presented in the Results section, we can conclude that providing the users with extra systematic training on the device had a highly positive effect on user experience.

Observed Problems

We observed a number of recurring problems encountered by the older adult users in each group. These problems occurred with more frequency in Group 1, which resulted in poorer usability metric outcomes. We grouped these problems into 3 categories—touch sensitivity, touch quality or accuracy, and user interface feedback. Within the first category, the users encountered problems when they held buttons for too long which would either deactivate the button press or initiate a secondary undesired function of the button. For example, holding a keypad letter for too long would input a number or symbols instead of letters. A related problem was observed when users were scrolling through a menu, such as in the settings page. Heavy touches while scrolling resulted in the user unintentionally entering an option (Multimedia Appendix 4). The user would then quickly become lost as they would not recognize the screen they were now presented with.

Users unfamiliar with touch screens had a tendency to leave their touch finger hovering near the screen when they were not interacting with the screen, causing unintentional and sometimes unnoticed screen presses. Unintentional touches also occurred if the user gripped the phone incorrectly. Problems within the second category, touch quality or accuracy, included inaccurate button striking and poor quality striking. For example, there

was tendency to aim too low when attempting to strike a button, usually owing to the angle at which the user held the screen (Multimedia Appendix 5). This, at times, led to excessive tapping, where the user would rapidly tap the screen in the hope of hitting the button correctly, leading them to unintentionally press a button in close proximity or to inadvertently press a button on the next screen. Finally, inadequate or unrecognized feedback on screen caused problems for some users. For example, many touch screen elements do not look like traditional buttons with clearly marked borders. This caused problems with striking accuracy. Sometimes users did not recognize the subtle changes in color or form that indicated a button had been successfully pressed. Other buttons had strange shapes which the user did not recognize (Multimedia Appendix 5), such as a triangle or an arrow for a send button, which led to hesitation and confusion.

Comments on Usability Metrics

The regression analysis performed on the usability metrics and how they affected the overall usability outcome from the SUS show that increased errors made and cues required were related to lower overall SUS scores. Increased ASQ scores (indicating greater task satisfaction) were related to higher SUS scores. Task completion times showed a weak to moderate relationship with SUS scores. This could be explained by the fact that many of the required tasks were not time intensive and had a very linear path toward the desired goal. Participants could not really "get lost" within a task or deviate too much from the optimum task path. Most participants who took longer to complete tasks did so because they were simply progressing at that pace. Therefore, task completion time may be a good indicator of overall usability and may depend on the context of the interaction and tasks involved.

With regard to the relationship between technology adoption predictors such as age, technology experience, and PU/PEoU on usability outcomes, we found no strong link for age and technology experience category although there was a moderate link for PU and PEoU.

Limitations and Recommendations Going Forward

Our methodology meant that the smartphone-specific training was maintained in parallel with the first 3 days of WIISEL training for Group 2 (see Table 3). There is a concern that this concurrent exposure, rather than implementing a cut-off where the smartphone-specific training ceased before Group 2 began using the WIISEL app, affected the outcome of the WIISEL usability data for Group 2 owing to the group having more total smartphone exposure time during the WIISEL app exposure. However, this approach was chosen such that the participant could achieve the benefit of the training they received on days −2 and −1, by recalling what they had learned from day 0 onward and receiving the appropriate guidance to optimize the training [33]. It is in this context that the participant achieves the benefit of the extra training, thereby improving their user experience and increasing their PEoU with the WIISEL system. We understand that for this training approach to be further operationalized, it may need to be streamlined or indeed undergo some structural changes. However, given the context of the study, we felt that our approach worked for the groups in question and allowed us to properly observe and measure their progress with the smartphone and the WIISEL app.

Regarding the specific problems we observed, we can make some instructional recommendations based on our training experience. These are presented in Textbox 1 and are presented within the categories of problems we identified in section Observed Problems.

Further to the specific recommendations for training older adults outlined in Textbox 1, we can make some more general recommendations. First, when approaching the older adult population, it is important to consider any cultural resistance and concerns that could act as hindrance for the uptake of technology-enabled health care. In particular, security, intrusiveness, lack of control, confidentiality, and usability issues can lead to a lack of trust in such technologies. We can assume that some cases of technology rejection are because of a lack of proper, scaffolded (scaffolded learning of technology is the use of progressive steps to allow the learner to gain independence in their use of the technology), and systematic introduction to the new technology [34,33]. This planned exposure, which should come in the form of appropriate training, must achieve the desired positive impact within a short window in order for technology acceptance rather than technology rejection to occur. The training must not only overcome apprehensions about how difficult the technology is to use but must also overcome any misconceptions about how useful the technology is. Older adults, as with many other user groups, overwhelmingly reject technology when there is unclear evidence of personal benefit or improvement in quality of life [35-37]. A key strategy of the training is to build and increase the user's perception of and trust in the technology. Although instructional activities can take the form of written materials, computer-based programs, or face-to-face communications, our experience in this study shows the benefit of short, yet intensive, periods of task-based learning with direct corrective feedback. This approach may have applications in other domains such as mobile learning in education [38].

Textbox 1. Recommendations for introducing smartphones to older adults.

Category and Solution or Recommendation

Touch sensitivity

Instruct the user to

- Touch and not press the screen. Instruct the user to strike deliberately and with the pad of the fingertip, not the finger nail (Multimedia Appendix 6).
- Carry out slow deliberate scrolls rather than quick stabbing motions.
- Remove fingers completely from the screen area when not interacting and teach them to hold the phone by the rails rather than with the digits wrapped around.
- When moving the textbox cursor, use light, slow, deliberate movement of the index finger.

Touch quality or accuracy

Instruct the user to

- Aim for the top portion of the button when striking (Multimedia Appendix 7).
- Hold the phone parallel to their eye line.
- Say the word "smartphone" (or similarly long word) after a button press before attempting to press it again if a button does not respond immediately.
- Avoid touching the tops or bottoms of the screen when swiping.
- Avoid fingers of the holding hand coming in contact with the touch screen surface.

User interface feedback

Teach users about

- The different types of touch elements on the screen which may not look like traditional buttons. For example, triangles or arrows for sending messages.
- Teach users to recognize the subtle feedback signals used to indicate button presses, such as slight color changes and button jumps.
- Teach users to recognize "wait" feedback such as tail chasers, loading bars, and hourglasses.

Conclusions

In this paper, we have discussed significant findings from a usability design study where a training intervention was developed and tested to introduce elderly users with limited or novice technology experience to a smartphone-based connected health care system. Our findings show the importance of properly introducing—in a scaffolded and systematic fashion—older adults to technology to improve their technology acceptance and enhance their user experience. Through our research in this specific context, we feel that it is possible to build or increase trust in connected, mobile, and wearable health technology through the design and deployment of meaningful instructional activities. Although there currently does not exist a structured training methodology for mHealth with older users, the methods and lessons learned in this paper can be used to conceptualize, design, and implement appropriate, bespoke training strategies. We understand that effort and time will not always be available to carry out training on a prolonged basis; however, scaffolded and structured intervention is necessary to ensure successful adoption of useful mHealth technology by elderly users. In addition to informing our future work, we hope the development of the WIISEL system and the guidelines and geragogical (geragogy is a theory which argues that older adults are sufficiently different that they warrant a separate educational theory) activities enumerated here will be widely used for those designing and developing connected health devices and infrastructure for older adults. More importantly, the involvement of end users is a key strategy for recognizing and removing barriers and mitigating design limitations, but our research has shown that this must be carefully planned to influence, drive, and refine systematically the iterative development of connected, mobile, and wearable health care technologies. We think this paper provides a useful enumerative approach to plan and conduct usability evaluations of smartphone apps and to gather user experience validation data, particularly in the domain of education and learning.

Acknowledgments

This work was partly funded by the EU FP7 project WIISEL (Wireless Insole for Independent and Safe Elderly Living; project number FP7-ICT-2011-288878).

Authors' Contributions

The experiment was conceived by RH, GOL, and LRQ. Observations were carried out by RH, and the data were analyzed by RH, GOL, and LRQ. All data and findings were reviewed by TH, LG, and ARM, who also provided methodological support. Recruitment and ethical application support was provided by TS.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Flowchart showing how each participant is classified in terms of mobile technology capability.

[PNG File, 132KB - humanfactors_v5i2e18_app1.png]

Multimedia Appendix 2

WIISEL user manual.

[PNG File, 568KB - humanfactors_v5i2e18_app2.png]

Multimedia Appendix 3

Phone user manual.

[PNG File, 442KB - humanfactors_v5i2e18_app3.png]

Multimedia Appendix 4

(A) The user is scrolling down to access the WIISEL option in the menu; (B) The user is tapping the screen too hard while scrolling and unknowingly presses the Earth app option (the white dot indicates the strike location); (C) The user is now suddenly in an unfamiliar screen and is at risk of pressing further buttons within this option as their hand may still be carrying out a scrolling action.

[PNG File, 1MB - humanfactors_v5i2e18_app4.png]

Multimedia Appendix 5

(A) The user attempts to strike the “Done” button but aims too low (white dot indicates the location of the strike); (B) During smartphone training, there were times when users felt that feedback on the screen was not appropriate; in this example, the user struggles to find the send button to send the text message (the horizontal triangle).

[PNG File, 1MB - humanfactors_v5i2e18_app5.png]

Multimedia Appendix 6

Finger form when touching actuation areas.

[PNG File, 697KB - humanfactors_v5i2e18_app6.png]

Multimedia Appendix 7

Striking accuracy.

[PNG File, 514KB - humanfactors_v5i2e18_app7.png]

References

1. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
2. Wac K. Smartphone as a personal, pervasive health informatics services platform: literature review. Yearb Med Inform 2012;7:83-93. [Medline: 22890347]
3. Smith A. Pewinternet. 2014. Older Adults and Technology Use URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2018-04-12] [WebCite Cache ID 6ycOGDPS0]

4. O'Brien MA, Rogers WA, Fisk AD. Understanding age and technology experience differences in use of prior knowledge for everyday technology interactions. ACM Trans Access Comput 2012 Mar 01;4(2):1-27. [doi: 10.1145/2141943.2141947]
5. Gerber T, Olazabal V, Brown K, Pablos-Mendez A. An agenda for action on global e-health. Health Aff (Millwood) 2010 Feb;29(2):233-236. [doi: 10.1377/hlthaff.2009.0934] [Medline: 20348066]
6. Silva PA, Holden K, Jordan P. Towards a List of Heuristics to Evaluate Smartphone Apps Targeted at Older Adults. 2015 Mar 30 Presented at: 48th Hawaii International Conference on System Sciences (HICSS); January 5-8, 2015; Kauai, Hawaii p. 3237-3246.
7. 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 01;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
8. Harte RP, Glynn LG, Broderick BJ, Rodriguez-Molinero A, Baker PM, McGuiness B, et al. Human centred design considerations for connected health devices for the older adult. J Pers Med 2014 Jun 04;4(2):245-281 [FREE Full text] [doi: 10.3390/jpm4020245] [Medline: 25563225]
9. Dickinson A, Newell AF, Smith MJ, Hill RL. Introducing the Internet to the over-60s: developing an email system for older novice computer users. Interact Comput 2005 Dec;17(6):621-642. [doi: 10.1016/j.intcom.2005.09.003]
10. Welch S, Kim SJ. Determining the effect of menu element size on usability of mobile applications. In: Marcus A, editor. Design, User Experience, and Usability. Web, Mobile, and Product Design. DUXU 2013. Lecture Notes in Computer Science, vol 8015. Berlin, Heidelberg: Springer; 2013.
11. Castilla D, Garcia-Palacios A, Bretón-López J, Miralles I, Baños RM, Etchemendy E, et al. Process of design and usability evaluation of a telepsychology web and virtual reality system for the elderly: Butler. Int J Hum Comput Stud 2013 Mar;71(3):350-362. [doi: 10.1016/j.ijhcs.2012.10.017]
12. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178 [FREE Full text]
13. Stara V, Harte R, Di Rosa M, Rossi L, ÓLaighin G. Toward a connected health system for older adults: lessons learned. In: Goonetilleke R, Karwowski W, editors. Advances in Physical Ergonomics and Human Factors. Advances in Intelligent Systems and Computing. Cham: Springer; 2016:857.
14. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too!Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013 May 06;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
15. European Commission. Developing an insole for elderly fall prevention - new European project WIISEL URL: https://ec.europa.eu/digital-single-market/en/news/developing-insole-elderly-fall-prevention-new-european-project-wiisel [accessed 2018-04-12] [WebCite Cache ID 6ycRk7Me1]
16. CDC Injury Center. CDC. Important Facts about Falls URL: https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html [accessed 2018-04-12] [WebCite Cache ID 6ycS2zbqH]
17. Nielsen Norman Group. Nngroup. Thinking aloud: The #1 usability tool URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ [accessed 2018-04-12] [WebCite Cache ID 6ycSBpROb]
18. Oulasvirta A, Wahlström M, Anders Ericsson K. What does it mean to be good at using a mobile device? An investigation of three levels of experience and skill. Int J Hum Comput Stud 2011 Mar;69(3):155-169. [doi: 10.1016/j.ijhcs.2010.11.003]
19. Blackler A. Intuitive Interaction with Complex Artefacts: Empirically-Based Research. Germany: VDM Verlag Dr. Müller; 2008.
20. Ryu MH, Kim S, Lee E. Understanding the factors affecting online elderly user's participation in video UCC services. Comput Human Behav 2009 May;25(3):619-632. [doi: 10.1016/j.chb.2008.08.013]
21. Cheung R, Vogel D. Predicting user acceptance of collaborative technologies: an extension of the technology acceptance model for e-learning. Comput Educ 2013 Apr;63:160-175. [doi: 10.1016/j.compedu.2012.12.003]
22. Davis FD. Perceived usefulness, perceived ease of use,and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
23. Svoboda E, Richards B, Leach L, Mertens V. PDA and smartphone use by individuals with moderate-to-severe memory impairment: application of a theory-driven training programme. Neuropsychol Rehabil 2012;22(3):408-427. [doi: 10.1080/09602011.2011.652498] [Medline: 22292945]
24. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
25. Fry H, Ketteridge S, Marshall S. A Handbook for Teaching and Learning in Higher Education Enhancing Academic Practice. New York: Routledge; 2009.
26. Lewis JR. Psychometric evaluation of an after-scenario questionnaire for computer usability studies. SIGCHI Bull 1990 Dec 01;23(1):78-81. [doi: 10.1145/122672.122692]
27. Cooke L. Assessing concurrent think-aloud protocol as a usability test method: a technical communication approach. IEEE Trans Profess Commun 2010 Sep;53(3):202-215. [doi: 10.1109/TPC.2010.2052859]

28. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Minh Tri Do T, Blom J, Gatica-Perez D. Smartphone usage in the wild: a large-scale analysis of applications and context. 2011 Nov 14 Presented at: ICMI '11 Proceedings of the 13th international conference on multimodal interfaces; November 14-18, 2011; Alicante, Spain.
30. Karikoski J, Soikkeli T. Contextual usage patterns in smartphone communication services. Pers Ubiquit Comput 2011 Dec 30;17(3):491-502. [doi: 10.1007/s00779-011-0503-0]
31. Lewis JR, Sauro J. Quantifying the User Experience: Practical Statistics for User Research. Amsterdam: Elsevier; 2012.
32. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
33. Hickman JM, Rogers WA, Fisk AD. Training older adults to use new technology. J Gerontol B Psychol Sci Soc Sci 2007 Jun;62 Spec No 1:77-84. [Medline: 17565168]
34. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478.
35. Gitlin LN. Why older people accept or reject assistive technology. J Am Soc Aging 1995;19(1):41 [FREE Full text]
36. Rama Murthy S, Mani M. Discerning rejection of technology. SAGE Open 2013 Apr 23;3(2):215824401348524 [FREE Full text] [doi: 10.1177/2158244013485248]
37. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
38. Baran E. A review of research on mobile learning in teacher education. J Educ Technol Soc 2014;17(4):17-32 [FREE Full text]

Abbreviations

ASQ: After Scenario Questionnaire
PEoU: perceived ease of use
PU: perceived usefulness
SUS: System Usability Scale
WIISEL: Wireless Insole for Independent and Safe Elderly Living

Edited by G Eysenbach; submitted 19.03.17; peer-reviewed by F Marvel, R Berenbaum, J Garcia Marin; comments to author 27.04.17; revised version received 21.07.17; accepted 18.09.17; published 26.04.18.

Please cite as:
Harte R, Hall T, Glynn L, Rodríguez-Molinero A, Scharf T, Quinlan LR, ÓLaighin G
Enhancing Home Health Mobile Phone App Usability Through General Smartphone Training: Usability and Learnability Case Study
JMIR Hum Factors 2018;5(2):e18
URL: http://humanfactors.jmir.org/2018/2/e18/
doi:10.2196/humanfactors.7718
PMID:29699969

Original Paper

Development of a Just-in-Time Adaptive mHealth Intervention for Insomnia: Usability Study

I Wayan Pulantara[1], PhD; Bambang Parmanto[1], PhD; Anne Germain[2], PhD

[1]Health and Rehabilitation Informatics Laboratory, Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA, United States

[2]Sleep and Chronobiology Laboratories, Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Bambang Parmanto, PhD
Health and Rehabilitation Informatics Laboratory
Department of Health Information Management
University of Pittsburgh
6026 Forbes Tower
Pittsburgh, PA, 15260
United States
Phone: 1 412 383 6649
Email: parmanto@pitt.edu

Abstract

Background: Healthy sleep is a fundamental component of physical and brain health. Insomnia, however, is a prevalent sleep disorder that compromises functioning, productivity, and health. Therefore, developing efficient treatment delivery methods for insomnia can have significant societal and personal health impacts. Cognitive behavioral therapy for insomnia (CBTI) is the recommended first-line treatment of insomnia but access is currently limited for patients, since treatment must occur in specialty sleep clinics, which suffer from an insufficient number of trained clinicians. Smartphone-based interventions offer a promising means for improving the delivery of CBTI. Furthermore, novel features such as real-time monitoring and assessment, personalization, dynamic adaptations of the intervention, and context awareness can enhance treatment personalization and effectiveness, and reduce associated costs. Ultimately, this “Just in Time Adaptive Intervention” for insomnia—an intervention approach that is acceptable to patients and clinicians, and is based on mobile health (mHealth) platform and tools—can significantly improve patient access and clinician delivery of evidence-based insomnia treatments.

Objective: This study aims to develop and assess the usability of a Just in Time Adaptive Intervention application platform called iREST (“interactive Resilience Enhancing Sleep Tactics”) for use in behavioral insomnia interventions. iREST can be used by both patients and clinicians.

Methods: The development of iREST was based on the Iterative and Incremental Development software development model. Requirement analysis was based on the case study’s description, workflow and needs, clinician inputs, and a previously conducted BBTI military study/implementation of the Just in Time Adaptive Intervention architecture. To evaluate the usability of the iREST mHealth tool, a pilot usability study was conducted. Additionally, this study explores the feasibility of using an off-the-shelf wearable device to supplement the subjective assessment of patient sleep patterns.

Results: The iREST app was developed from the mobile logical architecture of Just in Time Adaptive Intervention. It consists of a cross-platform smartphone app, a clinician portal, and secure 2-way communications platform between the app and the portal. The usability study comprised 19 Active Duty Service Members and Veterans between the ages of 18 and 60. Descriptive statistics based on in-app questionnaires indicate that on average, 12 (mean 12.23, SD 8.96) unique devices accessed the clinician portal per day for more than two years, while the app was rated as “highly usable”, achieving a mean System Usability Score score of 85.74 (SD 12.37), which translates to an adjective rating of “Excellent”. The participants also gave high scores on “ease of use and learnability” with an average score of 4.33 (SD 0.65) on a scale of 1 to 5.

Conclusions: iREST provides a feasible platform for the implementation of Just in Time Adaptive Intervention in mHealth-based and remote intervention settings. The system was rated highly usable and its cross-platformness made it readily implemented within the heavily segregated smartphone market. The use of wearables to track sleep is promising; yet the accuracy of this technology needs further improvement. Ultimately, iREST demonstrates that mHealth-based Just in Time Adaptive Intervention is not only feasible, but also works effectively.

(JMIR Hum Factors 2018;5(2):e21) doi:10.2196/humanfactors.8905

KEYWORDS

Just-in-Time Adaptive Intervention; JITAI; mobile health; mHealth; sleep; insomnia; usability; smartphone; iREST

Introduction

Insomnia is a prominent sleep problem. Defined as "a difficulty in falling asleep, difficulty staying asleep, and non-restorative sleep," [1] insomnia can contribute to further symptoms upon waking such as fatigue, impaired concentration, and mood disturbance [2]. Approximately 30% of adults in the United States have at least one of the symptoms of insomnia [3]. Diagnostic rates of insomnia however—based on criteria from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* —range between 5%-20% in the general adult population [4] and 20%-30% in primary care medical settings [5,6].

Since insomnia poses serious mental and physical health hazards, developing more efficacious treatment options is imperative. In general, there are two types of treatment for insomnia: pharmacological and behavioral. Hypnotic agents such as benzodiazepine receptor agonist (BZRA) drugs, are widely available, easy to use, and have rapid and sustained efficacy [7]. BZRA and other pharmacological treatments however, may lead to dependence and substance abuse [8]. Recently, studies have found that behavioral treatments such as Cognitive Behavioral Therapy for Insomnia (CBTI) [9] can be as effective as pharmacological treatments [10]. Furthermore, these behavioral treatments are often preferred by patients [11] and have been shown to have both short-term and long-term efficacy [12,13] with few apparent adverse effects.

Still, despite evidence that sleep disturbances are a modifiable threat to psychological and physical health, the use of evidence-based behavioral sleep treatments remains limited. By far the most limiting factor in making CBTI widely available is the shortage of trained clinicians. Although CBTI typically lasts for only eight sessions, a licensed psychologist trained in behavioral psychology must conduct these sessions [6]. Such restrictions pose an impediment to providing CBTI since there is currently a critical shortage of clinicians trained in evaluating and effectively treating sleep disturbances using behavioral strategies. Furthermore, there is also a geographic barrier to care: trained behavioral sleep clinicians are concentrated mostly in just a few major cities, while the need for them is dispersed throughout the entire country, especially in rural areas.

In response to these challenges, several internet-based solutions have emerged. SHUTi (SHUTi, Charlottesville, VA), as an example, provides CBTI through a web-based application. The results of this intervention are promising [14,15]. The SHUTi platform however requires that the user adhere to taking subjective assessments (eg, sleep diaries and quizzes), reading the modules, and watching provided videos. Moreover, SHUTi works on predetermined "if-then" algorithms, thus limiting the option of tailoring the intervention to a wide variety of individual response-to-treatment patterns, environmental contexts (eg, working schedule, daily routine), and condition severities. The rapid adoption of mobile technologies such as smartphones and wearable devices may help in the development of remote behavioral interventions for insomnia. For example, smartphone-based interventions may outperform web-based approaches such as SHUTi because they allow for continuous, prospective offline use, sensor access and integration, two-way messaging between patients and clinicians, and context awareness. Furthermore, wearable and smartphone-based sensors allow for the objective measurement of sleep and wake patterns instead of relying only on bias-prone, subjective sleep diaries.

Current mobile technologies can further tailor insomnia intervention by dynamically adapting both the assessment and intervention. For example, adaptability can be achieved by allowing the clinician to change the amount or the interval of sleep restriction prescribed to a patient in response to changes in the patient's sleep pattern and working hours. This sort of adaptability requires personalization of the intervention not only at the beginning of the episode of care, but also frequent iterative adjustments during care. When this adaptive intervention is combined with a smartphone-based approach, the result is the "Just in Time Adaptive Intervention" (JITAI) model [16].

An increasing number of studies have been conducted to assess the effect of JITAI on regulating human health behavior [17-21]. No generalizable application platform however, is yet available for JITAI. Such a platform would include ready-to-use, cross-platform, and reusable components like libraries, communication platforms, sensor integration, database access, and a logical infrastructure. The platform would allow application developers to customize the architecture to support a variety of health behavioral change interventions without having to build the system from scratch.

This work therefore aims to develop a JITAI application platform for an implementation in behavioral sleep interventions. This ostensible application has been called "interactive Resilience Enhancing Sleep Tactics," or iREST.

Methods

Preliminary Works

In the Health and Rehabilitation Informatics Laboratory at University of Pittsburgh, we have developed a JITAI platform—that is, a generalized platform conducive for work on a variety of health-intervention cases such as depression, anxiety disorders, smoking cessation, weight management, chronic condition management, and insomnia [22-25]. From a technological perspective, the difference between implementation across these health-intervention cases occurs mainly in the content of data, information presentation, and data collection methods (wearable devices used); however, the communication infrastructure and the service-oriented architecture remain constant across all cases.

Study Design

Phase 1: Development of iREST System

In this study, we developed the iREST system based on our JITAI platform in accordance with the needs of a behavioral sleep intervention study. In implementing the JITAI platform into the iREST system, we have followed the Iterative and Incremental Development (IID) software development model [26].

Requirement analysis was performed based on traditional behavioral insomnia workflows, clinician inputs, as well as a previous Brief Behavioral Therapy for Insomnia (BBTI) (a shorter, but equally effective version of CBTI), a military study, [27] and a previous implementation of the JITAI platform [24]. This analysis process was focused on determining the "context" and the "contents" of a JITAI delivery system for behavioral sleep intervention. A context is the type of health or behavioral therapy on which JITAI is implemented, for example: child anxiety, depression, insomnia, smoking cessation, and weight management. Contents, on the other hand, comprise such things as assessments, education materials, and guidelines needed to be communicated to achieve the goals of each context.

JITAI's requirement (Multimedia Appendix 1) for self-administered measurements has necessitated the development of numerous metrics in the iREST system. In iREST, users must fill out an electronic sleep log [28], a weekly assessment regimen, the Patient Health Questionnaire [29], the Asberg Rating Scale for side effects [30], and the User Global Impression of Improvement [31]. Additionally, clinicians must be able to access and complete the Clinician Global Impression [32] on a weekly basis so as to chart the user's ostensible progress.

Moreover, in using the JITAI platform to develop an effective iREST application, other functional requirements such as reminders and notifications, multimedia education and information delivery, real-time communication, and automatic data collection must be considered. These functional requirements have necessitated an integration of technologies like push-notifications, a secure messaging system, and wearable or Fitbit (Fitbit, San Francisco, CA) interface respectively.

Several of JITAI's non-functional requirements must also be accounted for and implemented during the development of the iREST app. These non-functional requirements include: privacy and security implementations, cross-platform capability, access and distribution concerns, safeguards for assurance and reliability, and a method for maintaining a separation of concerns. Addressing these non-functional requirements necessitates integrating an encryption technology, offering the app on various digital marketplaces like Google Play (Google, Mountain View, CA) and the Apple App Store (Apple, San Jose, CA), and developing a recovery procedure in the event of service disruption to minimize downtime.

Phase 2: Usability Evaluation

A pilot study was conducted to evaluate the usability of the iREST mobile health (mHealth) tool. The purpose of the usability study was to reveal how real patients and clinicians interact with the iREST system, gather their feedback, and improve the system based on the results.

The University's Institutional Review Board approved the present study. Participant recruitment and screening were conducted by the University of Pittsburgh Military Sleep Tactics and Resilience Research Team. Active Duty Service Members (ADSM) and Veterans between the ages of 18 and 60 were recruited through postcard, flyer, study website, social media and Facebook (San Francisco, CA) and public television. Since the study is a "Bring Your Own Device" (BYOD) study, to be eligible, ADSM and Veterans had to own a smartphone with internet access and be fluent in the use of a smartphone. Other eligibility criteria included:

1. Endorsing significant sleep complaints as determined by a baseline score higher than 5 on the Pittsburgh Sleep Quality Index (PSQI) [33];
2. Having a baseline score greater than 10 on the Insomnia Severity Index (ISI) [34]
3. Having sleep complaints for at least 1 month.

Exclusion criteria included:

1. A history of psychotic disorder or bipolar disorder
2. Suspected or previous diagnosis of sleep apnea narcolepsy or other sleep disorder requiring further evaluation and treatment;
3. Severe or untreated psychiatric disorder associated with marked impairments in functioning
4. Being pregnant or lactating
5. A scheduled/imminent military deployment during the study.

During the first office visit, participants were required to participate in a tutorial on how to use the iREST app. After this first visit, participants could try the app for 7-10 days. Following that period however, participants were required to return to the office to complete a "first impression" usability questionnaire as well as to provide feedback about the app in general. Afterwards, participants were instructed to continue using the app for the next 4-6 weeks of their BBTI. After this 4-to-6–week period, participants again returned to the research office for a postintervention usability assessment using the same assessment tools as those performed during the return visit (first impression). Those usability questionnaires implement in the study were the System Usability Scale (SUS) [35] and a modified version of Telerehabilitation Usability Questionnaire (TUQ) [36]. The SUS is a simple, ten-item scale giving a global view of subjective assessments of usability, while the TUQ—which is currently undergoing validation—measures several usability factors, including usefulness, usability, effectiveness, reliability, and satisfaction. Twenty-one questions were derived from previously validated questionnaires, including the Technology Acceptance Model, Perceived Usefulness and Ease of Use, the Telemedicine Satisfaction Questionnaire, and the Post-Study System Usability Questionnaire/Computer System Usability Questionnaire. In addition to formative usability questionnaires, participants were also asked to provide quantitative feedback or comments about the use of the iREST app. Measuring the system usability twice, before and after intervention, allowed for observation of whether habituation

affects participant perception of the system's ease-of-use. We hypothesized that habituation would not significantly affect usability in a negative way. Therefore, a paired Student t-test was performed to compare the SUS and the TUQ scores, preintervention and postintervention.

Adherence was calculated as half of the total participant logs (half, because there are two logs each day: wake log and sleep log) over the total number of days that participants used the app in the study. The completion time for each log was calculated by measuring the time lapse between the moment in which participants began accessing the sleep/wake log screen and the moment in which they hit the save button (for example, completed the logs). The calculation is performed automatically by the iREST system. In addition, the overall usage was estimated by calculating the number of unique devices with an iREST app accessing the iREST server per day.

Phase 3: Wearable Sensor Integration and Evaluation

In addition to the usability evaluation, we also evaluated the feasibility of further improving participant experience by using wearable sensors, which can potentially remove the burden of entering sleep diary data manually. To explore this potential, seven Fitbit Charge wristbands were randomly assigned to participants. Participants assigned with Fitbit bands were required to wear the band to measure their sleep patterns in addition to filling out the in-app sleep diary. After the BBTI intervention, sleep diary and Fitbit-reported sleep data were compared. This was meant to measure the degree of agreement between sleep parameters reported subjectively by the participants and measured by Fitbit devices: a high degree of agreement would indicate a higher potential for Fitbit to replace the need of manual sleep diary entries in our future BBTI interventions.

IBM (Armonk, NY) SPSS Statistics software version 24.0 was used for data analysis for Fitbit vs sleep diary comparison, while GraphPad (La Jolla, CA) Prism version 7 was used to build Bland-Altman plots. Sleep diary and Fitbit-reported sleep data were compared. First, intraclass correlation coefficients ($ICC_{2,1}$) were used to examine agreement between sleep parameters taken from the Fitbit and sleep diary data. An ICC ≥0.75 was considered excellent, 0.60–0.74 good, 0.40–0.59 fair and<0.40 poor [37]. A Bland-Altman plot [38] was used to visualize any systematic difference between values reported by the two measurements.

Results

Development Results

Currently, the iREST system (Figure 1) consists of a cross-platform smartphone app, a clinician portal, and a secure 2-way communication platform that connects the app and the portal.

Mobile Application (App) Features

The iREST mobile app (Figure 2) is used by the patient to record sleep data, present feedback and related education materials, and provide cues and notifications. Following are the app's main features:

Wake Log and Sleep Log

The wake log and sleep log of the iREST app is an electronic adaptation of the Pittsburgh Sleep Diary. The wake log records users' daytime activities that may impact healthy sleep practices. Such effecters include caffeine and alcohol consumption, number and duration of daytime naps, and exercise events. Users are intended to complete the wake log right before going to bed. Conversely, the sleep log tracks users' sleep parameters—including sleep latency, number and duration of wake-up after sleep onset episodes, bedtime, and wake time—dreams or nightmares, and perceived sleep quality. Users are required to fill out the sleep log immediately upon awakening to reduce recall bias.

For both sleep and wake logs, the app records time-stamps at the commencement and completion of each entry. Furthermore, the logs implement "validation checking", a function meant to ensure the thorough completion of app tasks. For example, while the user completes the logs, validation checking immediately alert the user if they have made any mistakes or missed any fields. Moreover, validation checking uses the previous entry as a default value for each new entry to reduce time and user burden in filling out the logs.

Weekly Assessment

The weekly assessment is a regimen of assessments administered to users on a weekly basis. These assessments consist of the Generalized Anxiety Disorder scale 2-items [39], the Patient Health Questionnaire-2 [29], the Patient-Rated Global Impression of Improvement [31], and a modified Asberg Rating Scale [30] side effect questionnaire that measure both the users' weekly progress and any potential side effects from treatment. This assessment appears to the user only when the clinician schedules it.

Sleep Education and Personalized Sleep Tips

Sleep education contains information about sleep, insomnia, brain sleep-mechanisms, and healthy sleep practices. These educational materials are always available on the iREST app. Additionally, the personalized sleep tips offer specific information on how to address or overcome certain behaviors, cognitions, or events (like nightmare episodes) that may be perceived as barriers to healthy sleep. Clinicians can prescribe sleep tips based on reports from users' sleep/wake logs. For example, if a user reports having a nightmare, the clinician can prescribe tips aimed at "getting rid of bothersome dreams" directly to the user's iREST app.

Secure Messaging

Secure messaging allows real-time message exchange between clinicians and users while maintaining high privacy and security—two factors that are often lacking on regular text messaging and short message services. Multiple security measures are implemented in the secure massaging feature, including a strong protocol for communication between the iREST app and server using Transport Layer Security, a secure encryption key exchange, an encoding and enciphering of messages, and achieving an encrypted database behind a firewall.

The secure messaging feature allows users and clinicians to exchange information that may not be readily available through the app's other functionalities. For example, through secure messaging—after reading the personalized sleep tips prescribed by the clinician—a user may request additional information on specific sleep problems. The clinician can then reply with links to additional resources.

App Dashboard

The iREST app's dashboard provides "at-a-glance" views of key performance indicators on individual treatment progress. In sleep interventions, these indicators can be sleep parameters such as sleep efficiency, sleep latency, wake-up-after-sleep-onset, and total sleep time. The dashboard also contains indicators of logs and assessment completion. In addition, the dashboard provides visual notification for new messages and new tips that are received from the clinician portal.

Clinician Portal Features

Like the iREST mobile app, the clinician portal (Figure 3) is an implementation of the portal portion of the JITAI application architecture, based on the requirements needed by the clinicians to fully support the intervention (as also described in Multimedia Appendix 1). Below are the main features available in the iREST clinician portal:

Clinician Dashboard

The iREST portal's dashboard provides data visualization of users' progress in the intervention, the intervention status as whole, and general views of the mHealth utilization. It allows clinicians to make priorities on resource allocation based on the severity of users' conditions. For example, users who frequently express sleep problems will have more clinician time than users whose interventions are going well.

Calendar and Scheduling

The calendar view allows clinicians to quickly assess the status of scheduled intervention components such as prescribed wake time and assessment schedules. This page also provides users with a mobile device status (active, idle, inactive) and shortcuts for creating new schedules for sending secure messages to users.

Intervention Prescription

Intervention prescription is the main feature for managing and prescribing intervention components. It provides users' daily sleep logs and weekly progress summaries. Based on these summaries, the portal suggests appropriate sleep prescriptions, and the clinicians then make judgments on which course of action to take, or which intervention components to prescribe to the users' mobile app.

Figure 1. A model representing the iREST app and clinician-portal's two-way interactions which include: assessment, education/information delivery, progress reporting, scheduling, notification delivery, and secure messaging.

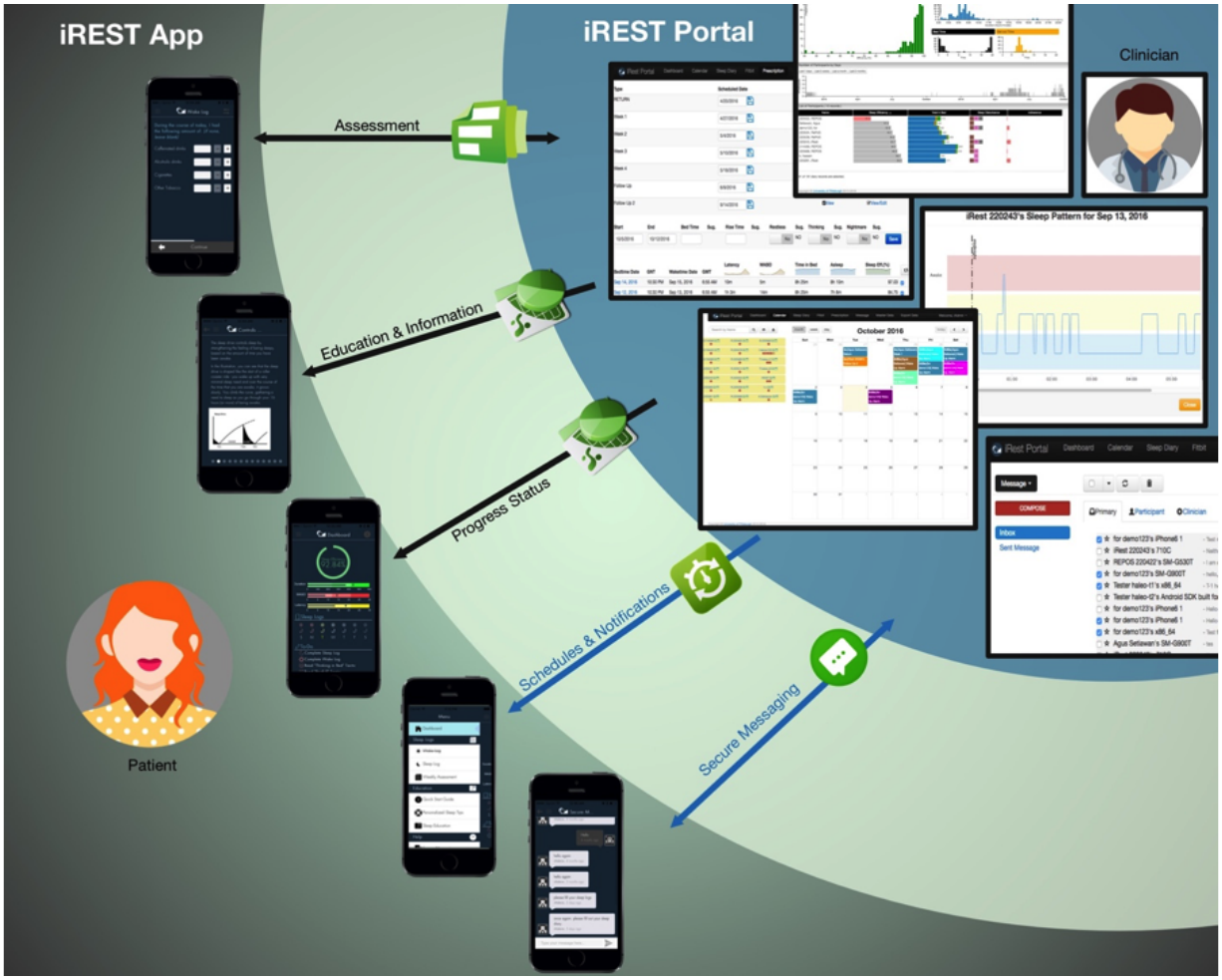
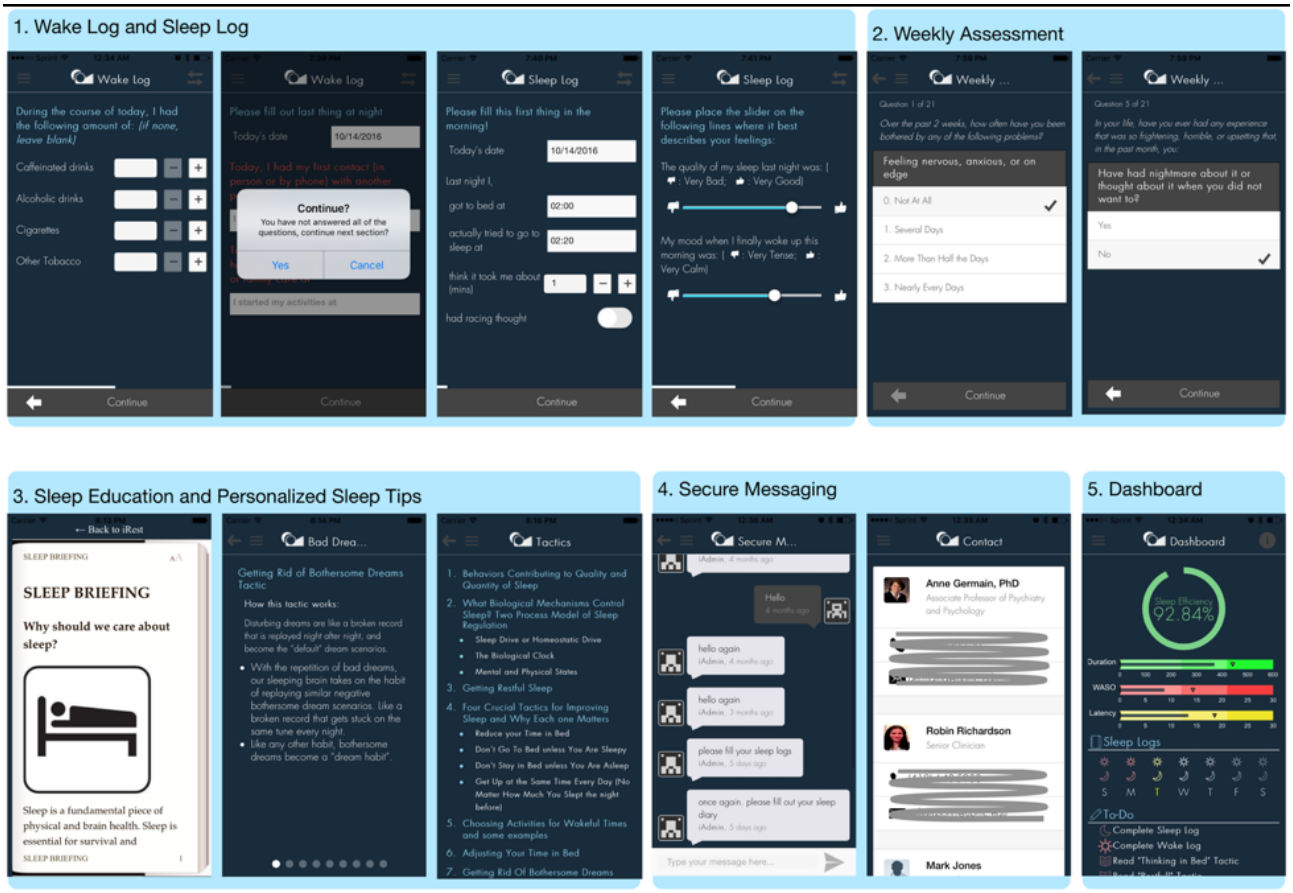


Figure 2. Screenshots of the various features implemented in the iREST app. 1) Wake activity and sleep pattern recording part of the iREST app, participants need to enter each section once a day; 2) Weekly assessment was scheduled each week to measure participants' clinical progress; 3) Snapshots of various sleep education and personalized tactics that the app provides based on each participant's condition; 4) Secure massaging feature that allows participants and clinicians to exchange information securely, eg, instead of using short message service/text. 5) The iREST app's dashboard, showing an at-a-glance view of individual participant current status and progress.



Wearable Sensor Integration

With the current clinician portal, only integration with Fitbit is supported. The integration functionality provides interfaces to perform sleep data imports from the Fitbit server.

Participants

As shown in Figure 4, a total of 99 individuals contacted the research program to inquire about the study, all expressing interest in participating. During the scripted telephone screening, 35 (36%) individuals did not respond after several attempts to contact them. Twelve individuals (19%) were found not eligible after telephone screening. ADSM and Veterans who passed the telephone screening attended the in-office diagnostic evaluation; ten individuals (19%) were excluded in this phase. Twenty-nine ADSM and Veterans provided written informed consent; however, seven of them (24%) withdrew from the study before the intervention. Out of 22 who started the intervention, nineteen (19) participants (86%) completed posttreatment and follow-up assessment. Six (32%) participants used an iPhone or iOS device, and the other 13 (68%) used an Android device.

Descriptive statistics were performed to describe the demographic characteristics of the study participants using frequencies for categorical variables, means, and standard deviations for continuously measured demographic variables. Demographic information obtained at baseline is provided in Table 1.

Usage Characteristics

One way to describe the overall usage of the system is by calculating the number of unique devices accessing the iREST portal per day. As seen in Figure 5, on average, there were at least 12 (mean 12.23, SD 8.96) unique devices accessing the portal daily for more than two years following the iREST study commencement. On the app side, according to the Apple App Store and Google Play statistics from September 2016, the app was downloaded and installed 247 times (182 on Android [Google, Mountain View, CA] and 67 iOS). The number of downloads was significantly higher than the number of participants in the study (in total only 29 participants downloads), which may indicate that there was high demand for a sleep or insomnia app on the market. Currently, the app is active on 53 devices (47 Android and 6 iOS). In addition to the current study, the iREST mHealth system was also used to support at least two other sleep research studies.

Figure 3. Screenshots from the iREST clinician portal. 1) The portal's calendar view which simplify the way clinicians manage participants' scheduling; 2) The prescription window where clinicians can view individual participant's status, view treatment suggestions calculated by the system, and prescribe appropriate intervention; 3) An example of a participant's sleep pattern retrieved from Fitbit; 4) Secure messaging on the clinician side; 5) The clinician's dashboard where a clinician can view the whole status of participants under their care, see the clinical indicators/signs of progress and prioritize treatments.

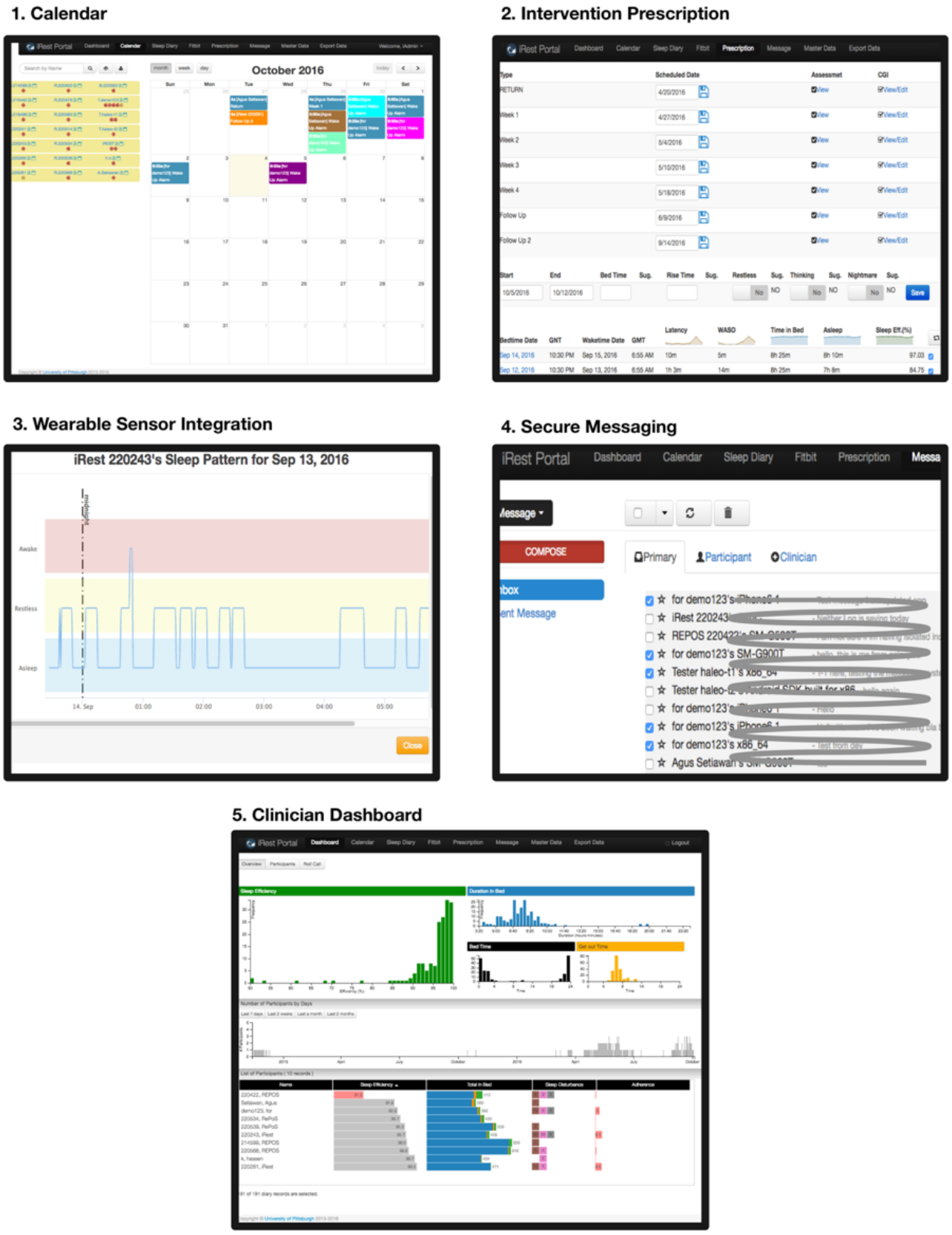


Figure 4. iREST Usability Study's Participant Flow. WD: withdrawn.

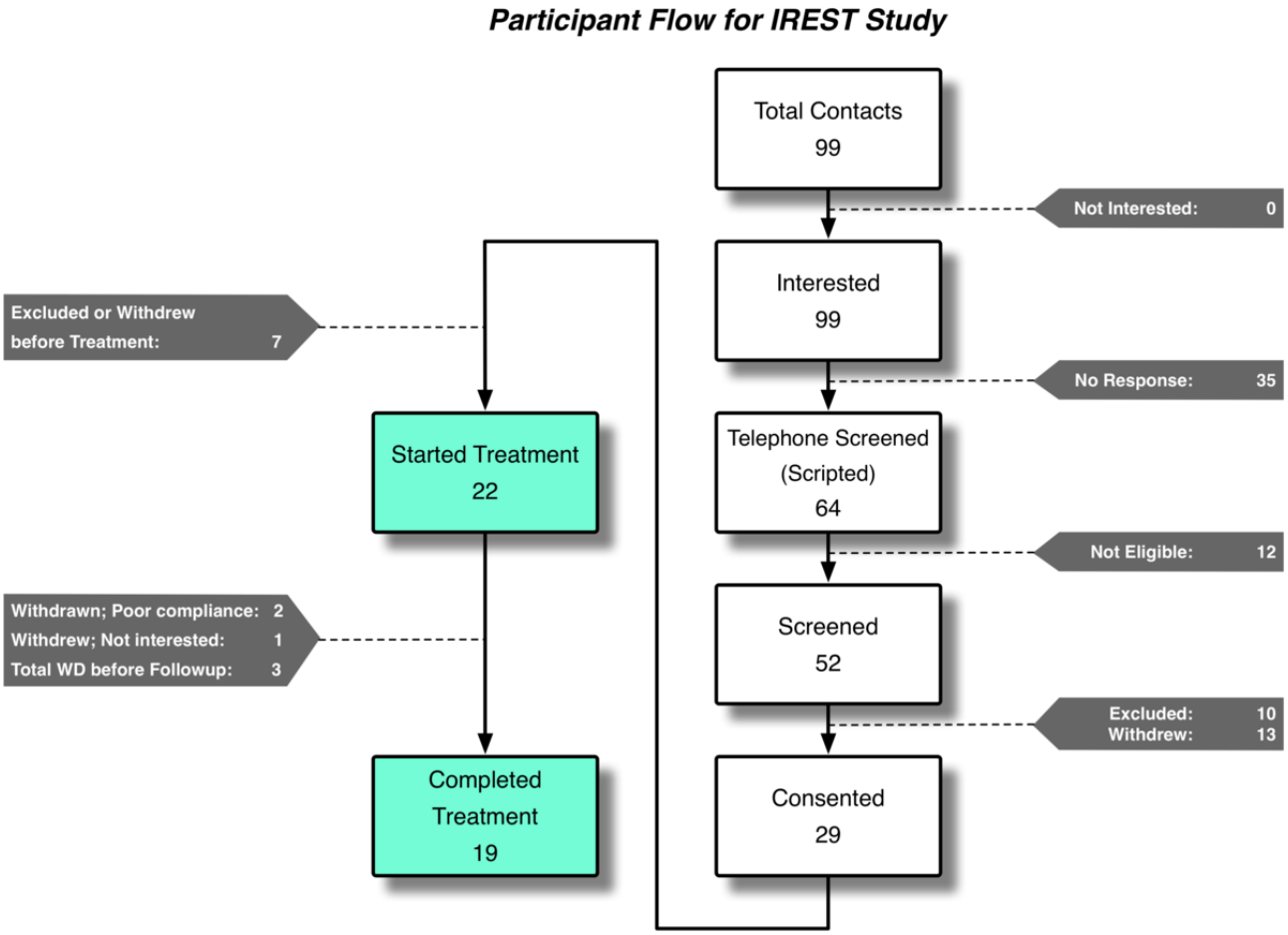


Table 1. Demographic and clinical baseline.

Variable	Value
Male, n (%)	18 (82)
Caucasian, n (%)	17 (77)
Age, mean (SD)	38.7 (9.7)
Army, n (%)	14 (64)
Pittsburgh Sleep Quality Index, mean (SD)	11.9 (3.9)
Insomnia Severity Index, mean (SD)	17.4 (4.0)
Epworth Sleepiness Scale, mean (SD)	7.4 (4.6)

During the study, our online server experienced an unplanned outage resulting in no data collection over a three-day period for two participants. This unexpected problem was subsequently fixed by making the system capable of handling server and connection outages. Even with this outage, on average participants completed 91.11% of the required twice-a-day sleep diary entries, only failing to fill out less than 3 days' worth of sleep diaries throughout the course of the study. This adherence percentage is significantly higher than the average technology-mediated insomnia treatment adherence of 52% [40]. It took an average of less than two minutes (108.53 seconds, SD 26.19) to complete each assessment.

Usability Results

In the postintervention follow-up visit, 17 out of 19 participants finished the poststudy questionnaires (the SUS and the TUQ). The sample size is considered appropriate according to the Problem Discovery Rate Model, which is widely used to serve in formative usability evaluations [41,42]. According to the model, 85% of usability problems were revealed using five participants, and almost 100% of problems using 14 participants [43]. The participants rated the app as highly usable with a mean SUS score of 85.74 (SD 12.37), which translates to adjective ratings of *"Excellent"* [44]. On the TUQ, participants were satisfied with the iREST app and would consider using it in the future (average score of 4.31 out of 5, SD 0.63). They also gave high scores on "ease of use and learnability" with an average

score of 4.33 (SD 0.65). In assessing room for improvement, the sections for "interface quality" and "reliability" received slightly lower scores, although still above average, with a mean score of 4.05 (SD 0.85) and 3.88 (SD 0.70), respectively. Server outage may have contributed to lower scores on reliability, while the lower interface quality score shows the need for more meaningful data visualization and better overall user-interface design.

When compared with pretreatment scores, both SUS and TUQ posttreatment scores were higher. The results show that participants continued to rate the iREST app as highly usable even as they became more familiar with the system; in other words, rather than fostering *contempt*, in this case, familiarity can be said to *breed contentment*. Furthermore, the improvement in the interface quality score on the TUQ was statistically significant (Table 2), and a noticeable score increase was observed on Reliability (mean increase of 0.45). Continued user-centered improvements (eg, incorporating users' feedback and addressing user interface, UI, interaction problems) in user interface and system reliability most likely contributed to the noticeable increase in TUQ scores for those two areas.

On the qualitative usability assessment, participants provide individual comments and feedback about the app. Responses were generally categorized into five types:

1. General comments about the app
2. Comments about the graphical user interface and navigation
3. Comments about the sleep logging process
4. Comments about sleep education features
5. Questions and problem reporting.

Participants expressed liking the application generally with reported comments such as: *"[I] like the front page a lot, [I] find it useful and attractive"*; *"[The app is] very easy to navigate, [I] found that the data uploaded quickly"*; *"[I] like the morning reminder to fill out wake time diary."* Participants also pointed out issues and made suggestions, such as: *"[The app's format for time input was tedious, [this] needs improvement"*; *"[I was] frustrated by [the] text overlap [that occurs] when [the] device is held horizontally"*; *"[Developers should] have the SE% graphic replaced by something, eg, tracking how many logs were entered on time."*Multimedia Appendix 2 contains a portion of the reported feedback and comments. This feedback was used to iteratively improve the iREST app and as input for future developments.

Additionally, qualitative analysis of participants' feedbacks on iREST app identified several potential improvements and additional features suggestions. Most of the critical suggestions have already been addressed and incorporated into the app during the development iterations. Some suggested improvements however remain to be addressed in future developments. These include:

1. An informative and concise, but customizable data visualization on the app's dashboard
2. A smart data input, in which the app learns from previously-entered data about each participant's usual sleep habits to reduce participant burden
3. Include general UI components
4. Implementing more reliable notifications and reminders
5. Streamlining the Fitbit integration

Figure 5. Statistical representation of daily unique device access to the iREST portal. Each line represents number of unique participants accessing the portal for each day.

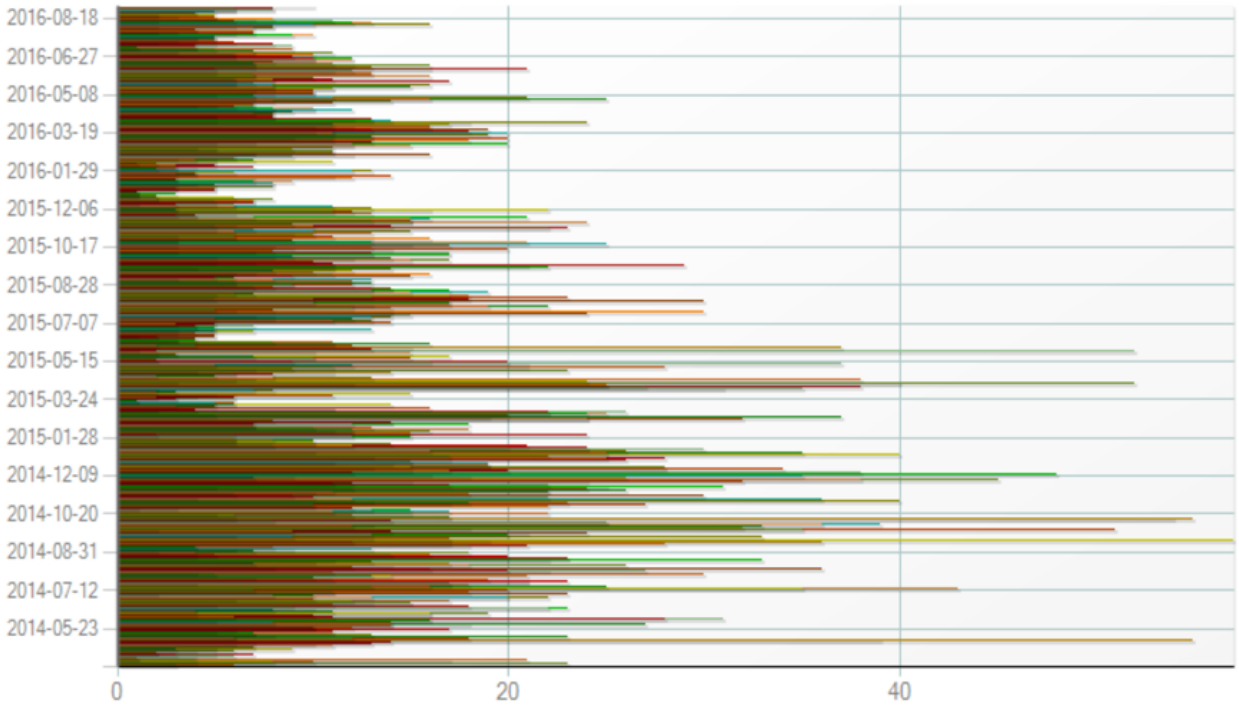


Table 2. A paired *t* test comparison of Usability Questionnaires Scores (System Usability Scale and Telerehabilitation Usability Questionnaire) between pretreatment and posttreatment. SUS: System Usability Scale; TUQ: Telerehabilitation Usability Questionnaire.

Variable	Before treatment (SD)	After treatment (SD)	Pretreatment to postmean score change		
			Estimates (SE)	*t* statistics (*df*)	*P* value
SUS	78.04 (13.66)	85.74 (12.37)	6.61 (3.65)	1.81 (13)	.09
TUQ					
Ease of use	4.15 (0.55)	4.33 (0.65)	0.14 (0.18)	0.76 (10)	.47
Interface Quality	3.65 (0.65)	4.05 (0.85)	0.55 (0.17)	3.25 (10)	.009[a]
Interaction Quality	3.77 (0.76)	3.95 (0.78)	0.20 (0.14)	1.11 (10)	.29
Reliability	3.58 (0.53)	3.88 (0.70)	0.45 (0.22)	2.09 (10)	.06
Overall Satisfaction	3.96 (0.80)	4.31 (0.63)	0.36 (0.22)	1.62 (10)	.14

[a]Statistically significant.

Table 3. Fitbit versus iREST Sleep Diary.

Variables	Fitbit, mean (SD)	iREST Diary, mean (SD)	Mean difference (SE)	*P* value	$ICC_{2,1}$ (range)
Sleep onset latency (minutes)	0.38 (0.38)	18.60 (1.63)	18.23 (1.66)	<.001	0.15 (-0.123-0.153)
Wakefulness after sleep onset (minutes)	5.64 (0.58)	20.19 (1.97)	14.54 (1.86)	<.001	0.174 (0.038-0.305)
Total in bed (minutes)	422.9 (6.48)	430.9 (6.00)	8.08 (4.80)	.09	0.705 (0.628-0.768)
Total sleep time (minutes)	416 (6.39)	392.14 (6.82)	–24.69 (4.79)	<.001	0.737 (0.667-0.794)
Sleep efficiency (%)	98.59 (0.16)	90.53 (0.74)	-8.06 (0.70)	<.001	0.144 (0.006-0.276)
Good night/fall asleep time	11:42:12PM (5.55)	11:27:06PM (5.35)	–15.10 (3.95)	<.001	0.738 (0.668-0.795)
Good morning/awake time	06:45:00AM (6.87)	06:38:00AM (6.25)	–7.03 (4.39)	.11	0.777 (0.715-0.826)

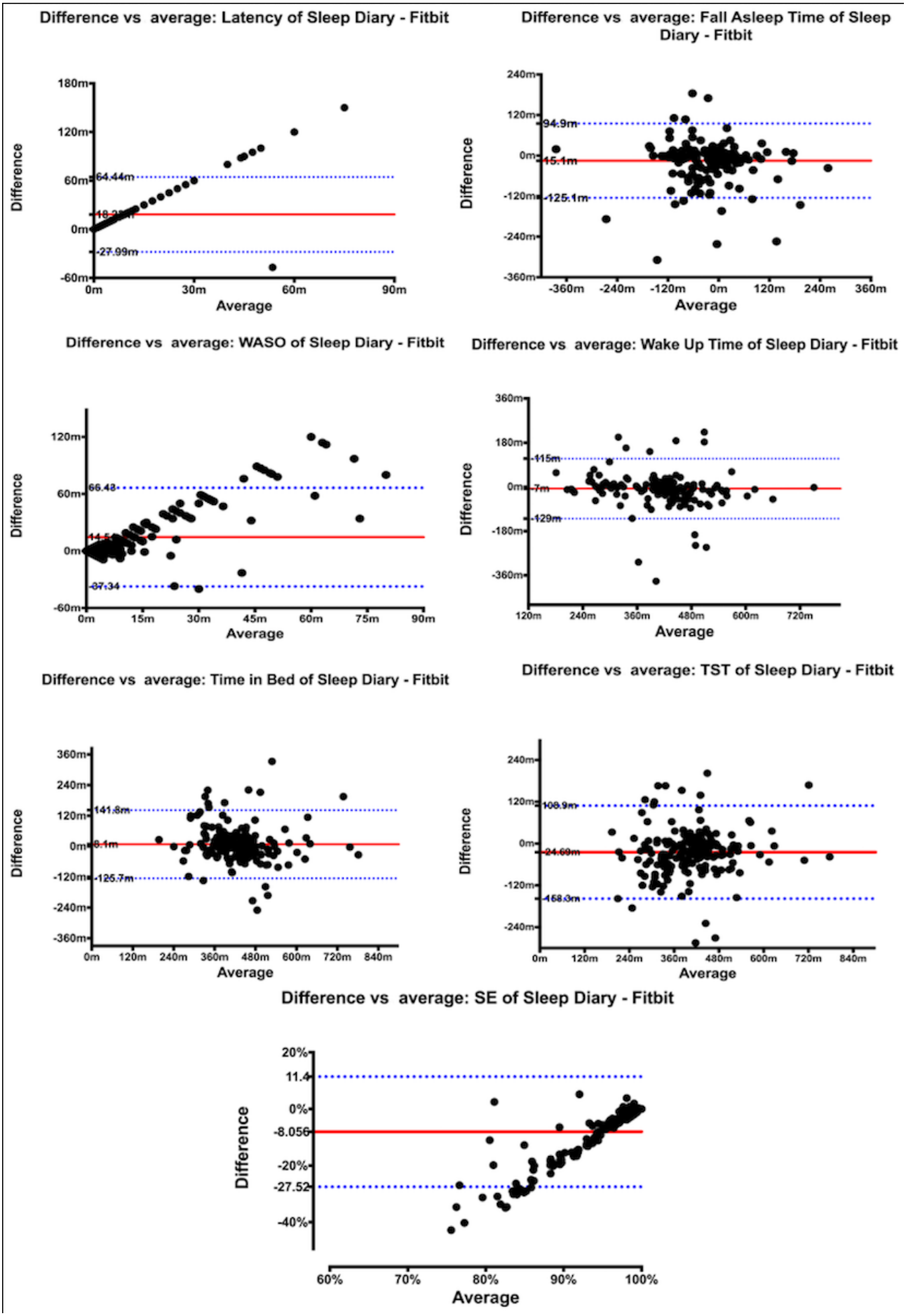
Fitbit Integration

Seven participants were assigned with Fitbit Charge throughout the course of the usability study. In total, 202 paired (Fitbit vs sleep diary) nights were acquired. We utilized the automatic sleep detection feature available on the Charge model, in which the wristband automatically detects when the wearer falls asleep and wakes up without manual input (eg, pressing a button). This feature, although convenient for participants, greatly underestimates latency to onset of the first sleep epoch (sleep onset latency, SOL). As a result, Fitbit only reported one instance of SOL (SOL>0 minute) out of 202-recorded nights. Due to limitations in statistical analysis packages, variables representing clock time, such as Good Night Time (GNT) and Good Morning Time (GMT) are translated into minutes-distant-from-midnight (12:00 AM). For example, 11:00 PM on GNT was translated into “–60” (60 minutes before midnight), and 5:15 AM in GMT was translated into “315” (315 minutes after midnight).

As shown in Table 3, significant statistical differences were found for the following variables recorded between Fitbit and sleep diaries: latency (SOL), wakefulness after sleep onset (WASO), total sleep time (TST), GNT, and sleep efficiency. With diaries recording longer means of latency, longer means of WASO, shorter means of TST, earlier GNTs and smaller average sleep efficiencies. No significant differences however were found on total in bed (TIB; longer in sleep diary) and GMT (earlier in sleep diary). Furthermore, good intraclass correlations were observed for TST ($ICC_{2,1}$=0.737, *P*<.001), GNT ($ICC_{2,1}$=0.738, *P*<.001), and TIB ($ICC_{2,1}$=0.705, *P*<.001). There was excellent agreement on GMT between Fitbit and sleep diary entries, with $ICC_{2,1}$=0.777, *P*<.001.

Bland and Altman difference plots (Figure 6), for these sleep parameters showed no statistically significant agreement between Fitbit and sleep diaries. As demonstrated in ICC analysis, the plots also show a higher level of disagreement between Fitbit and diaries for SOL, WASO and sleep efficiency. Moreover, proportional bias was observed for these three variables, and the disagreement between measurement modalities increased as the average value of SOL and WASO increased, and as the average value of sleep efficiency decreased from 100. It was observed that when an individual’s day-to-day sleep pattern variability increased, the level of agreement between Fitbit and sleep diaries decreased for that individual when compared with the group mean. For example, a participant whose WASO changed significantly from night to night (eg, from 0 on the first day, to 45 on the second day, and back to 15 on the third) is likely to have worse agreement on the Fitbit vs sleep diary WASO when compared with the rest of the sample.

Figure 6. Bland-Altman plots of Sleep Diary vs. Fitbit. SE: sleep efficiency; TST: total sleep time; WASO: wakefulness after sleep onset.



Discussion

Principal Findings

The JITAI application architecture used in iREST gives potential leverage for intervention-scientists in implementing mobile app solutions for other JITAI based projects. The architecture provides a wide variety of functionalities, design patterns, and guidelines that are readily implemented in various JITAI mobile app solutions. Also, the JITAI application architecture is cross-platform and therefore allows rapid deployment to various mutually incompatible mobile operating systems and opens the possibility for a BYOD approach, a feature that greatly increases the scalability of, and access to, interventions.

The usability evaluation of iREST showed that the app is highly usable and supports high adherence to treatment regimens. In addition, the evaluation allowed the detection of potential improvement of the iREST system based on participants' feedback and comments about the system during the usability study. These improvements have been incorporated in subsequent development iterations of iREST system.

The usability results demonstrated that not only is the IREST app applicable to implementing BBTI in a military population but is also usable and well received. Overall, participants were satisfied with the iREST application, finding it easy to use. All nineteen participants used the app daily to record their sleep with very few missed entries (on average less than 3) over the course of the 4-to-6-week intervention. The iREST app's sleep and wake logs were optimized for touch-based input, which reduced fill-out time and participants' burden.

Consistent with previous studies conducted in the use of movement-based sensors (eg, Fitbit, actigraphy) for measuring sleep [45–47], we found poor agreement between Fitbit and participants' reported sleep diaries, although clinically, the Fitbit data may be sufficiently used as a consideration for BBTI sleep prescriptions (eg, mean differences for sleep parameters between Fitbit and sleep diary are below the clinically significant threshold of 30 minutes). Improvements nevertheless need to be made to the architecture to increase the sensitivity of Fitbit's measuring of sleep parameters. As a possible solution for the next iteration of the iREST system, we plan to use a hybrid approach between the two modalities, to incorporate a machine learning algorithm and to allow participants to modify Fitbit reported data. Each modification will then feed into a machine-learning algorithm, so that the longer an individual uses the system, the less modification that will be needed to provide more accurate data.

Limitations

The present study was highly focused on patients' improvements and experiences in using the iREST system. The clinicians' perspectives however, are equally important. As mentioned before, the highly manual nature of BBTI supported by the easy-to-use iREST system is likely to facilitate the delivery of this treatment by mid-level (non-doctoral) clinicians; however, the present study has not provided a sufficient level of heterogeneity in a clinician sample to determine whether a comparable magnitude of improvement would also be observed with less experienced therapists. In the current study, two clinicians with extensive experience in behavioral sleep treatments administered the intervention.

Moreover, although this paper included a usability study and utilization analysis on the iREST patient app, no usability data from the clinician portal was explicitly reported. Clinician usability is something that might be investigated in future studies.

Currently, the iREST wearable integration only supports Fitbit devices. Furthermore, the iREST app does not support background services (ie, running in the background and without user intervention), which would allow the app to keep running after the smartphone screen is locked and the app minimized. This feature is important for real-time communication such as initiation of video or audio calls. Future development to implement this feature has been planned.

Conclusion

The iREST system provides a feasible platform for implementation of JITAI in mHealth-based and remote intervention settings. The use of Fitbit as an objective measure for ambulatory sleep pattern assessment showed promising results, yet further improvement is needed.

Ultimately, iREST demonstrates that mHealth-based JITAI model works effectively while achieving an excellent usability rating. Although the current implementation was only aimed at insomnia treatment, iREST has the potential to be deployed towards other behavioral health interventions. Furthermore, the promising results in the current pilot study, open the pathway for larger study on clinical feasibility of iREST intervention.

Acknowledgments

We thank Hassen Khan, Noelle Rode, Mark Jones and Robin Richardson for their help and support in iREST study and for their helpful feedback. We also like to thank Justin Stec for his contribution in writing this manuscript. This project was supported in part by grants from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant numbers 90RE5018 and 90DP0064). NIDILRR is a Center within the Administration for Community Living, Department of Health and Human Services.

Conflicts of Interest

None declared

Multimedia Appendix 1

iREST System Requirements.

[PDF File (Adobe PDF File), 35KB - humanfactors_v5i2e21_app1.pdf]

Multimedia Appendix 2

Participant Feedbacks and Comments on iREST App.

[PDF File (Adobe PDF File), 42KB - humanfactors_v5i2e21_app2.pdf]

References

1. Daroff RB. The International Classification of Sleep Disorders: Diagnostic and Coding Manual. Neurology 1991 Jan 01;41(1):160-160. [doi: 10.1212/WNL.41.1.160]
2. Roth T. Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med 2007 Aug 15;3(5 Suppl):S7-10 [FREE Full text] [Medline: 17824495]
3. Ancoli-Israel S, Roth T. Characteristics of insomnia in the United States: results of the 1991 National Sleep Foundation Survey. I. Sleep 1999 May 01;22 Suppl 2:S347-S353. [Medline: 10394606]
4. Ohayon M. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Medicine Reviews 2002 May;6(2):97-111. [doi: 10.1053/smrv.2002.0186]
5. Simon GE, VonKorff M. Prevalence, burden, and treatment of insomnia in primary care. Am J Psychiatry 1997;154(10):1417-1423.
6. Shochat T, Umphress J, Israel A. Insomnia in primary care patients. Sleep: Journal of Sleep Research & Sleep Medicine 1999;22(Suppl 2):S359-S365.
7. Krystal A, Erman M, Zammit G, Soubrane C, Roth T. Long-term efficacy and safety of zolpidem extended-release 12.5 mg, administered 3 to 7 nights per week for 24 weeks, in patients with chronic primary insomnia: a 6-month, randomized, double-blind, placebo-controlled, parallel-group, multicenter study. Sleep 2008 Jan 1;31(1):79-90. [doi: 10.1093/sleep/31.1.79]
8. Goodwin R, Hasin D. Sedative use and misuse in the United States. Addiction 2002 Apr 25;97:555-562. [doi: 10.1046/j.1360-0443.2002.00098.x]
9. Edinger JD, Wohlgemuth WK, Radtke RA, Marsh GR, Quillian RE. Cognitive Behavioral Therapy for Treatment of Chronic Primary Insomnia. JAMA 2001 Apr 11;285(14):1856. [doi: 10.1001/jama.285.14.1856]
10. Smith MT, Perlis M, Park A, Smith MS, Pennington J, Giles DE, et al. Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. Am J Psychiatry 2002 Jan;159(1):5-11. [doi: 10.1176/appi.ajp.159.1.5] [Medline: 11772681]
11. Morin C, Gaulier B, Barry T, Kowatch R. Patients' acceptance of psychological and pharmacological therapies for insomnia. Sleep 1992;15(4):302-305. [doi: 10.1093/sleep/15.4.302]
12. Irwin MR, Cole JC, Nicassio PM. Comparative meta-analysis of behavioral interventions for insomnia and their efficacy in middle-aged adults and in older adults 55+ years of age. Health Psychol 2006 Jan;25(1):3-14. [doi: 10.1037/0278-6133.25.1.3] [Medline: 16448292]
13. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
14. Thorndike FP, Ritterband LM, Gonder-Frederick LA, Lord HR, Ingersoll KS, Morin CM. A randomized controlled trial of an internet intervention for adults with insomnia: effects on comorbid psychological and fatigue symptoms. J Clin Psychol 2013 Oct;69(10):1078-1093 [FREE Full text] [doi: 10.1002/jclp.22032] [Medline: 24014057]
15. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an Internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
16. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
17. Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychol Addict Behav 2014 Sep;28(3):639-650. [doi: 10.1037/a0034747] [Medline: 25000269]
18. Consolvo S, McDonald D, Toscos T, Chen M, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of ubifit garden. 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 05 - 10, 2008; Florence, Italy p. 1797-1806.
19. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013 Apr;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]

20. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
21. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A Context-Sensing Mobile Phone App (Q Sense) for Smoking Cessation: A Mixed-Methods Study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]
22. Yu DX, Parmanto B, Dicianno BE, Pramana G. Accessibility of mHealth Self-Care Apps for Individuals with Spina Bifida. Perspect Health Inf Manag 2015;12:1h [FREE Full text] [Medline: 26755902]
23. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
24. Juengst SB, Graham KM, Pulantara IW, McCue M, Whyte EM, Dicianno BE, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015 Aug;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]
25. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: A Novel mHealth System for Supporting Self-Care in Management of Complex and Chronic Conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10 [FREE Full text] [doi: 10.2196/mhealth.2391] [Medline: 25100682]
26. Larman C, Basili V. Iterative and incremental developments. a brief history. Computer 2003 Jun;36(6):47-56. [doi: 10.1109/MC.2003.1204375]
27. Germain A, Richardson R, Stocker R, Mammen O, Hall M, Bramoweth AD, et al. Treatment for insomnia in combat-exposed OEF/OIF/OND military veterans: preliminary randomized controlled trial. Behav Res Ther 2014 Oct;61:78-88 [FREE Full text] [doi: 10.1016/j.brat.2014.07.016] [Medline: 25194223]
28. Monk T, Reynolds C, Kupfer D, Buysse D, Coble P, Hayes A, et al. The Pittsburgh Sleep Diary. Journal of Sleep Research 1994;3(2):111-120. [doi: 10.1111/j.1365-2869.1994]
29. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
30. Asberg M, Cronholm B, Sjoqvist F, Tuck D. Correlation of Subjective Side Effects with Plasma Concentrations of Nortriptyline. BMJ 1970 Oct 03;4(5726):18-21. [doi: 10.1136/bmj.4.5726.18]
31. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. American Journal of Obstetrics and Gynecology 2003 Jul;189(1):98-101. [doi: 10.1067/mob.2003.379]
32. Busner J, Targum S. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 2007 Jul;4(7):28-37 [FREE Full text] [Medline: 20526405]
33. Buysse D, Reynolds C, Monk T, Berman S, Kupfer D. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4]
34. Bastien C, Vallières A, Morin C. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Medicine 2001 Jul;2(4):297-307. [doi: 10.1016/S1389-9457(00)00065-4]
35. Sauro J. MeasuringU. 2011. Measuring Usability with the System Usability Scale (SUS): MeasuringU URL: https://measuringu.com/sus/ [accessed 2017-08-31] [WebCite Cache ID 6t8MwBLwH]
36. Schutte J, Gales S, Filippone A, Saptono A, Parmanto B, McCue M. Evaluation of a telerehabilitation system for community-based rehabilitation. Int J Telerehabil 2012 Jun;4(1):15-24 [FREE Full text] [doi: 10.5195/ijt.2012.6092] [Medline: 25945193]
37. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [Medline: 22833776]
38. Martin Bland J, Altman D. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8]
39. Skapinakis P. The 2-item Generalized Anxiety Disorder scale had high sensitivity and specificity for detecting GAD in primary care. Evid Based Med 2007 Oct;12(5):149. [doi: 10.1136/ebm.12.5.149] [Medline: 17909240]
40. Horsch C, Lancee J, Beun RJ, Neerincx MA, Brinkman W. Adherence to Technology-Mediated Insomnia Treatment: A Meta-Analysis, Interviews, and Focus Groups. J Med Internet Res 2015;17(9):e214 [FREE Full text] [doi: 10.2196/jmir.4115] [Medline: 26341671]
41. Lewis C, Rieman J. Testing The Design With Users. In: Task-centered user interface design: A Practical Introduction. Boulder: Clayton Lewis and John Rieman; 1993:77-96.
42. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. 1993 Presented at: the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; April 24 - 29, 1993; Amsterdam p. 206-213. [doi: 10.1145/169059.169166]
43. Lewis J. Sample sizes for usability tests. Interactions 2006 Nov 01;13(6):29. [doi: 10.1145/1167948.1167973]
44. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies 2009 May;4(3):114-123.

45. McCall C, McCall WV. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. J Sleep Res 2012 Feb;21(1):122-127 [FREE Full text] [doi: 10.1111/j.1365-2869.2011.00917.x] [Medline: 21447050]
46. Lockley SW, Skene DJ, Arendt J. Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. J Sleep Res 1999 Sep;8(3):175-183. [doi: 10.1046/j.1365-2869.1999.00155.x]
47. Buysse DJ, Cheng Y, Germain A, Moul DE, Franzen PL, Fletcher M, et al. Night-to-night sleep variability in older adults with and without chronic insomnia. Sleep Med 2010 Jan;11(1):56-64 [FREE Full text] [doi: 10.1016/j.sleep.2009.02.010] [Medline: 19962939]

Abbreviations

ADSM: Active Duty Service Members
BYOD: Bring Your Own Device
BBTI: Brief Behavioral Therapy for Insomnia
BZRA: benzodiazepine receptor agonist
CBTI: Cognitive Behavioral Therapy for Insomnia
DSM IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
GNT: Good Night Time
GMT: Good Morning Time
ICC: intraclass correlation
JITAI: Just in Time Adaptive Intervention
mHealth: mobile health
PSQI: Pittsburgh Sleep Quality Index
RCT: randomized controlled trial
SE: sleep efficiency
SOL: sleep onset latency
SUS: System Usability Scale
TIB: total in bed
TST: total sleep time
TUQ: Telerehabilitation Usability Questionnaire
UI: user interface
WASO: wakefulness after sleep onset

Edited by G Eysenbach; submitted 05.09.17; peer-reviewed by S Ng, S Chambers, M O'Reilly; comments to author 23.11.17; revised version received 09.02.18; accepted 28.02.18; published 17.05.18.

Please cite as:
Pulantara IW, Parmanto B, Germain A
Development of a Just-in-Time Adaptive mHealth Intervention for Insomnia: Usability Study
JMIR Hum Factors 2018;5(2):e21
URL: http://humanfactors.jmir.org/2018/2/e21/
doi:10.2196/humanfactors.8905
PMID:29773529

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org