Contents

Original Papers

A Novel Method for Digital Pain Assessment Using Abstract Animations: Human-Centered Design Approach (e27689)
Nema Rao, Sophy Perdomo, Charles Jonassaint. ............................................................... 2

Clinician Perspectives on Unmet Needs for Mobile Technology Among Hospitalists: Workflow Analysis Based on Semistructured Interviews (e28783)
April Savoy, Jason Saleem, Barry Barker, Himalaya Patel, Areeba Kara. .................................................. 18

Key Challenges and Opportunities for Cloud Technology in Health Care: Semistructured Interview Study (e31246)
Kathrin Cresswell, Andrés Domínguez Hernández, Robin Williams, Aziz Sheikh. ........................................... 28

Comparing International Experiences With Electronic Health Records Among Emergency Medicine Physicians in the United States and Norway: Semistructured Interview Study (e28762)
Gracie Garcia, Christopher Crenner. .................................................................................. 39

Combining Farmers’ Preferences With Evidence-Based Strategies to Prevent and Lower Farmers’ Distress: Co-design and Acceptability Testing of ifarmwell (e27631)
Kate Gunn, Gemma Skaczkowski, James Dollman, Andrew Vincent, Camille Short, Susan Brumby, Alison Barrett, Nathan Harrison, Deborah Turnbull. ................................................................. 53

Educators’ Perspectives on Integrating Technology Into Sexual Health Education: Implementation Study (e31381)
Martha Decker, Salish Harrison, Melissa Price, Abigail Gutmann-Gonzalez, Jennifer Yarger, Rachel Tenney. ................................................................. 76

Using Health Concept Surveying to Elicit Usable Evidence: Case Studies of a Novel Evaluation Methodology (e30474)
Alex Mariakakis, Ravi Karkar, Shwetak Patel, Julie Kientz, James Fogarty, Sean Munson. .................................................. 86

Purpose Formulation, Coalition Building, and Evidence Use in Public–Academic Partnerships: Web-Based Survey Study (e29288)
Christina Kang-Yi, Amy Page. .................................................................................. 107

An Interactive Voice Response System to Increase Physical Activity and Prevent Cancer in the Rural Alabama Black Belt: Design and Usability Study (e29494)
Mohanj Thirumalai, Nashira Brown, Soumya Niranjjan, Sh’Nese Townsend, Mary Powell, Whitney Neal, Erica Schleicher, Venkatadri Raparla, Robert Oster, Wendy Demark-Wahnefried, Dori Pekmezli. .................................................................................. 126
Original Paper

A Novel Method for Digital Pain Assessment Using Abstract Animations: Human-Centered Design Approach

Nema Rao¹, MDes; Sophy Perdomo², MS, PhD; Charles Jonassaint², MHS, PhD

¹Microsoft Inc, SharePoint Spaces, Seattle, WA, United States
²Center for Behavioral Health and Smart Technology, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Charles Jonassaint, MHS, PhD
Center for Behavioral Health and Smart Technology
Department of Medicine
University of Pittsburgh
230 McKee Pl Suite 600
Pittsburgh, PA, 15213
United States
Phone: 1 4125869850
Email: cjonassaint@pitt.edu

Abstract

Background: Patients with chronic pain face several challenges in using clinical tools to help them monitor, understand, and make meaningful decisions about their pain conditions. Our group previously presented data on Painimation, a novel electronic tool for communicating and assessing pain.

Objective: This paper describes the human-centered design and development approach (inspiration, ideation, and implementation) that led to the creation of Painimation.

Methods: We planned an iterative and cyclical development process that included stakeholder engagement and feedback from users. Stakeholders included patients with acute and chronic pain, health care providers, and design students. Target users were adults with acute or chronic pain who needed clinical assessment and tracking of the course of their pain over time. Phase I (inspiration) consisted of empathizing with users, understanding how patients experience pain, and identifying the barriers to accurately expressing and assessing pain. This phase involved understanding how patients communicate pain symptoms to providers, as well as defining limitations of current models of clinical pain assessment tools. In Phase II (ideate) we conceptualized and evaluated different approaches to expressing and assessing pain. The most promising concept was developed through an iterative process that involved end users and stakeholders. In Phase III (implementation), based on stakeholder feedback from initial designs and prototypes of abstract pain animations (painimations), we incorporated all concepts to test a minimally viable product, a fully functioning pain assessment app. We then gathered feedback through an agile development process and applied this feedback to finalizing a testable version of the app that could ultimately be used in a pain clinic.

Results: Engaging intended users and stakeholders in an iterative, human-centered design process identified 5 criteria that a pain assessment tool would need to meet to be effective in the medical setting. These criteria were used as guiding design principles to generate a series of pain assessment concept ideas. This human-centered approach generated 8 highly visual painimations that were found to be acceptable and useable for communicating pain with medical providers, by both patients with general pain and patients with sickle cell disease (SCD). While these initial steps continued refinement of the tool, further data are needed. Agile development will allow us to continue to incorporate precision medicine tools that are validated in the clinical research arena.

Conclusions: A multiphase, human-centered design approach successfully resulted in the development of an innovation that has potential to improve the quality of medical care, particularly for underserved populations. The use of Painimation may especially benefit the medical care of minority populations with chronic and difficult-to-treat pain, such as adults with SCD. The insights generated from this study can be applied to the development of patient-reported outcomes tools that are more patient-centered, engaging, and effective.

(JMIR Hum Factors 2022;9(1):e27689) doi:10.2196/27689

https://humanfactors.jmir.org/2022/1/e27689
KEYWORDS

pain; pain measurement; chronic pain; animations; mobile apps; human-centered design

Introduction

Background

Pain is the number one reason people access the health care system. The Centers for Disease Control and Prevention (CDC) reported that in 2019, approximately 20.4% of US adults had chronic pain, and 7.4% had high-impact, chronic pain. Similar statistics have been reported in Canada (18.9%) and Australia (17.9%), whereas in the United Kingdom the numbers are much higher (35%-51.3%) [1-4]. The cost of medical treatment and lost productivity due to pain exceeds US $635 billion each year in the United States, more than the cost of treating cardiovascular disease, cancer, or diabetes [5]. Chronic pain also significantly affects an individual’s quality of life, negatively impacting their ability to engage in day-to-day activities, and increasing risk for depression, anxiety, and opioid dependence [6,7].

Despite the significant impact of pain on population health outcomes, pain remains inadequately assessed in the health care setting [8]. Pain is a complex sensory and emotional experience that is often difficult to communicate [9]. Unidimensional pain measures, such as the numeric or visual analog pain scale, reduce the complex, multifaceted nature of the pain experience to a single number between 0 and 10 [10]. This oversimplification not only results in poor assessment of potential physiological mechanisms but also ignores the complex roles the patient’s thoughts and mood play in the patient’s pain experience [8].

In some subspecialty medical clinics, multidimensional measures are used, such as the McGill Pain Questionnaire [11], that attempt to take into account other facets of pain beyond intensity, such as pain location, quality, and affective response. However, these measures are often overly complex and rely on long lists of adjectives or esoteric phrases to describe pain that may alienate individuals with low literacy, individuals with dementia or other cognitive limitations, non-native English speakers, and many others with communication limitations [12,13]. With the current state of clinical pain assessment, even individuals without language limitations can have their needs misinterpreted, their symptoms ignored, or their credibility challenged [14]. Ineffective communication about pain may result in patient–clinician discordance, leading clinicians to intervene on poorly described and ill-defined targets, and patients to feel misunderstood and lose trust in their provider [15,16]. The inadequacy of pain assessment tools compromises medical providers’ ability to deliver quality care and improve clinical outcomes for their patients [10,17,18].

Painimation

To address the limitations of standard pain assessment, we used human-centered design methods to discover, design, and develop a novel method for assessing pain that leverages digital animations that we call painimations [19]. In this work, we hypothesized that an animation-based pain assessment tool would be more acceptable to patients with pain than traditional numerical and adjective-based pain assessments. Our work is particularly timely, given the recent promising evidence suggesting that digital health interventions are feasible, acceptable, and efficacious in a range of chronic medical conditions [20-25].

Our prior publication presented data comparing participants’ selection of painimations with their scores on validated, traditional pain scales that rely on pain adjectives and numerical scales [17]. This paper describes our process of using human-centered design to understand how patients experience and express their pain, how clinicians assess and diagnose pain, and how leveraging these observations led to the creation of a novel method for pain assessment: Painimation.

Our approach incorporated human-centered design principles, qualitative methods, and stakeholder engagement, and consisted of 3 distinct phases: the inspiration phase, the ideation phase, and the implementation phase [26,27]. After detailing the discovery and development process for a novel, animation-based pain assessment approach, we present initial user testing of the painimations, or abstract animations that can be visually configured to reflect pain quality, pattern, and intensity, as well as the overall Painimation prototype. Finally, we describe future directions for the use of Painimation and discuss how this digital animation approach has the potential to significantly improve medical assessment and treatment of acute and chronic pain.

Methods

Setting

The human-centered design process that resulted in the development of a Painimation prototype took place from January 2015 to May 2016. Key stakeholders were recruited from the Pittsburgh, Pennsylvania, metropolitan area and included patients with acute and chronic pain, clinicians, clinical researchers, and design students. All participants were 18 years of age or older. This project was approved by the University of Pittsburgh’s and Carnegie Mellon University’s Institutional Review Boards.

Phase I: Inspiration (Empathize, Understand, and Define)

Overview

Human-centered design is inherently an empathic process that attempts to set aside the investigators’ or designers’ assumptions about the world and gain insight into their users’ lived experience, perspectives, pain points, and needs [26]. The goal of Phase I was to empathize with the target user and understand how pain is experienced and communicated. The next step was to define the most prominent barriers to effective patient–provider communication, assessment, and treatment of pain in the health care setting. To accomplish this, we conducted one-on-one, in-depth, in-person interviews with patients with acute and chronic pain, clinicians, and researchers (Table 1).
Table 1. Questions from interviews using directed storytelling and modified think-aloud protocol.

<table>
<thead>
<tr>
<th>User and stakeholder</th>
<th>Clinician and clinical researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A successful</strong> experience I've had with a clinician around my pain assessment and management was SHORT STORY</td>
<td>The pain assessment protocol I follow is BRIEF OVERVIEW</td>
</tr>
<tr>
<td>I describe the pain communication between myself and my clinician as ADJECTIVE</td>
<td>Pain assessment is part of every interaction I have with a patient YES/NO</td>
</tr>
<tr>
<td>I summarize my clinician’s understanding of and assessment of my pain as ADJECTIVE</td>
<td>I use the following tools LIST/DESCRIBE</td>
</tr>
<tr>
<td>I describe my communication ability as ADJECTIVE</td>
<td>I document in the following way ADJECTIVE</td>
</tr>
<tr>
<td>I have been asked to rate my pain intensity on a scale like this YES/NO</td>
<td>(Numeric) pain scales are an effective/ineffective CHOOSE tool because REASON</td>
</tr>
<tr>
<td>The experience of using the scale was ADJECTIVE</td>
<td></td>
</tr>
<tr>
<td>During that interaction, I communicated the pain intensity that I felt YES/NO</td>
<td></td>
</tr>
</tbody>
</table>

**User and Stakeholder Interviews**

Interviews were conducted using directed storytelling [28], a design ethnography method, which allowed patients with a wide range of pain experiences to be interviewed, and yielded information about the contexts in which they had experienced pain as well as descriptions of successful and unsuccessful interactions with their clinicians.

The next part of the patient interviews consisted of a modified version of the think-aloud protocol [29], a method during which participants verbalize their thought process while doing specific tasks. The aim of this portion of the interview was to understand how patients think through 2 current pain scales: the Wong–Baker faces scale and the Numeric Rating Scale [18,30]. Additionally, patients were given a recall interview prompt to understand how they have used these scales in the past to describe their pain to medical providers.

**Clinician and Clinical Researchers Interviews**

As with the patient interviews, clinician interviews were conducted using directed storytelling to learn about their expertise and experiences in interacting with and treating patients with pain. All interviews were transcribed for later analysis.

**Phase II: Ideation (Generate Concepts and Designs)**

**Overview**

The goal of Phase II was to develop solutions to the problem defined in Phase I: how to best allow patients to express their pain and facilitate pain communication with health providers.

**Idea and Concept Development**

Analysis of the interviews from Phase I combined thematic analysis and the constant comparison method [31,32]. Codes were developed via open coding of the transcripts to determine topics and themes that emerged. Input from the designers/investigators on relevant topics was also integrated, resulting in a simultaneously inductive and deductive analysis process. Based on the topics identified in Phase I, we developed a set of criteria that needed to be met for a pain communication solution to be considered successful. These criteria served as design principles that guided the ideation stage where the designers generated a large number of concepts, or creative and innovative solutions to the pain communication problem.

Once several solutions, or concepts, are developed in an unrestricted brainstorm, all of the concepts are evaluated based on the design principles defined earlier. Any concepts that do not meet all of the design principles are discarded. The remaining concepts are ranked relative to 2 axes or factors: importance (ie, potential to impact the problem) and then difficulty (eg, cost, feasibility, scalability). Final concepts are selected based on their relative importance/difficulty and developed using generative storyboards to illustrate how the concepts might function in various scenarios. To test each concept, we conducted needs validation sessions, a design method for working with stakeholders to validate or disprove early ideas, to select a viable concept, and to transition it to the user evaluation stage.

**Painimation Drawing Exercises**

The process of developing the painimations began with the words used to describe the qualities of pain on the McGill Pain Questionnaire Short Form [11], a pain assessment method that measures pain intensity and quality using 15 descriptors of pain. Drawing exercises were conducted with a group of 16 design students from Carnegie Mellon, to develop visual depictions of the more commonly used pain adjectives. For this exercise, the design students were given a list of qualitative words that are currently used on the McGill Pain Questionnaire Short Form, such as stabbing, pounding, and shooting, and were asked to draw those words, creating a low or medium version, and a high version for each word. The selected words were those most frequently presented by patients and clinicians in Phase I exercises.

**Painimation Development**

Words from the McGill Pain Questionnaire Short Form [11] were clustered into a few groups, with the idea of creating painimations that would depict and represent different sensations. The first 3 types we explored were throbbing, shooting, and cramping. Deep and dull are terms that could be applied to other qualities, so these were clustered separately. Next, the visual variables that the painimations would represent
or communicate were listed. The final list included speed, saturation, focus, and size (Figure 1). Changing these variables would change the intensity of the pain depicted. These painimations were sent out to the design students in a survey with the question, “What kinds of pain do you believe these animations evoke?” The goal was to understand how participants would describe the qualities of these painimations, given the context of pain.

Figure 1. Visual considerations for painimations.

<table>
<thead>
<tr>
<th>Shape</th>
<th>Spheres to indicate sharpness</th>
<th>Lines (pain)</th>
<th>Mimicking electricity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus</td>
<td>In focus to communicate intensity</td>
<td>In focus to communicate intensity</td>
<td>In focus to communicate intensity</td>
</tr>
<tr>
<td>Color and saturation</td>
<td>Shades of red</td>
<td>Red and black for contrast</td>
<td>White instead of black (to cool more like electricity)</td>
</tr>
<tr>
<td>Movement and speed</td>
<td>Moving in a circle</td>
<td>Having a central point but shooting in all directions</td>
<td>Having a central point but shooting in all directions</td>
</tr>
<tr>
<td>Fluctuating or consistent</td>
<td>Fluctuating to indicate change over time</td>
<td>Fluctuating to indicate change over time</td>
<td>Fluctuating to indicate change over time</td>
</tr>
<tr>
<td>Size</td>
<td>Goes out of frame in certain places, to communicate intensity</td>
<td>Goes out of frame in certain places, to communicate intensity</td>
<td>Stays mostly within the frame</td>
</tr>
</tbody>
</table>

**Wireframe Creation**

A new set of 11 participants—patients with a history of pain, clinicians, and researchers—were recruited by word of mouth and asked to evaluate the painimations as well as the context of use through the think-aloud protocol. Basic wireframes for the pain assessment app were created to provide context for the painimations (Figure 2). The participants were asked, “How effective do you think this tool is in aiding your pain communication?” via a modified version of the think-aloud protocol. Similarly, clinicians and clinical researchers were asked “Would something like this work? Why or why not?”
Figure 2. Painimation wireframe early sketches.

Phase III: Implementation (Prototype, Test, and Iterate)

Overview
The goal of Phase III was to develop a minimally viable product to test with a small number of users. Once the painimations were refined based on user input from Phase II, we used an agile development process to build a fully functioning prototype of an app that utilized the painimations (i.e., Painimation). A final set of 8 painimations was developed and subjected to testing and further design iteration. The designers labeled the painimations based on what pain adjective the painimations were intended to represent. Two independent, graduate-level design students were asked to identify what pain type each painimation represented. Confirmation that the painimations approximated the pain adjective they were meant to represent would allow us to transition to pilot testing; otherwise, the painimations would go through another design iteration.

Pilot Testing Using a Case Patient Population: Adults With Sickle Cell Disease

The use-case scenario for Painimation was the assessment and treatment of sickle cell disease (SCD) pain. SCD is a genetic blood disorder that is characterized by unpredictable vaso-occlusive episodes that lead to severe acute pain often called “crisis” and can result in long-term organ damage, chronic pain, and other complications [33]. Patients living with SCD experience pain crisis as early as infancy, and the pain can transition to chronic pain during adolescence and young adulthood. Further, SCD primarily affects underserved, racial/ethnic minorities, and patients often experience discrimination in the medical system [34]. Thus, adults with SCD have long, many times difficult, historical experience with pain and communicating pain to medical providers; these conditions informed the development of this tool.

Participating adult patients with SCD and self-reported chronic pain were presented the 8 painimations and asked, “Would you find this animation applicable to your pain?” These patients were also asked about what types of pain they experience, how they track pain, and their history of pain communication interactions with providers.
Results

Phase I: Inspiration (Empathize, Understand, and Define)

User and Stakeholder Interviews

In total, 10 patients were interviewed, 6 with acute pain (mean age 42.5 years; range 25-50; 50% [n=3] female) and 4 with chronic pain (mean age 40.0 years; range 24-58; 75% [n=3] female). Participants with acute pain had experience with temporary bouts of pain lasting no more than a few days, and patients with chronic pain had a range of pain experiences all lasting more than 3 months. Participants with acute pain experienced a hairline fracture, kidney stones, a pulmonary embolism, postsurgery pain, a root canal, and a urinary tract infection, whereas those with chronic pain experienced migraines, fibromyalgia, vulvodynia, and chronic back pain. Participants with acute and chronic pain both reported having experience communicating pain with clinicians in the medical setting.

Directed storytelling interviews revealed that patients with acute and chronic pain both felt their exact pain was impossible to communicate due to its subjective nature and the individual response to it, both physical and mental. Patients with chronic pain expressed that they particularly struggled to find clinicians who knew and accepted their conditions.

Patients described communication about pain with their health provider as “successful” if they felt heard and understood. Likewise, pain communications were described as “unsuccessful” if there was a lack of understanding, feelings of being dismissed, or intimidated. Textbox 1 displays extracted quotes from these interviews.

The think-aloud protocol revealed that patients with chronic and acute pain both expressed some confusion around traditional pain scales because they felt these scales were “vague” and “ambiguous.” For example, several patients stated they had “no clue” what “worst possible pain” in the numerical pain scale meant.

Additionally, patients felt that traditional pain scales “lack specificity” and do not accommodate detailed answers. For example, on the numerical pain scale, one might want to say, “It’s an 8 when I am applying pressure, and a 7 when I am resting, and a 10 early in the morning.” Patients said that they used these scales to communicate their pain intensity because they had to; 5/10 respondents said their numerical pain rating did not feel accurate.

Textbox 1. Extracted descriptors for clinical communication.

<table>
<thead>
<tr>
<th>Successful</th>
<th>Unsuccessful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personable, friendly</td>
<td>Zero understanding</td>
</tr>
<tr>
<td>Professional</td>
<td>Accused me of lying</td>
</tr>
<tr>
<td>Dead-on</td>
<td>Impossible</td>
</tr>
<tr>
<td>Light at the end of a tunnel</td>
<td>Dismissive</td>
</tr>
<tr>
<td>Calming</td>
<td>Limited</td>
</tr>
<tr>
<td>Relieving</td>
<td>Intimidating</td>
</tr>
<tr>
<td>I felt in control</td>
<td>I felt stupid</td>
</tr>
<tr>
<td>I was actually being heard</td>
<td></td>
</tr>
<tr>
<td>Improved over time</td>
<td></td>
</tr>
</tbody>
</table>

Clinician and Clinical Researchers Interviews

A total of 7 individuals were interviewed, 4 clinicians (mean age 36.3 years; range 30-50 years, 50% [n=2] female) and 3 clinical researchers (mean age 50.7 years; range 36-58 years; 33% [n=1] female). Clinicians had experience in emergency medicine, general medicine, and physical therapy while clinical researchers had experience in clinical psychology, hematology, and anesthesiology. Clinicians had experience caring for patients with chronic and acute pain, while clinical researchers provided their clinical experience as well as a rich perspective into current research, challenges, and opportunities.
Directed storytelling interviews revealed clinicians’ and clinical researchers’ perspectives on traditional pain scales. Clinicians explained that a numeric value on the Numerical Pain Rating Scale is only meant to represent one person’s pain: “one person’s 5 can be compared to their 9, but you cannot compare two individuals’ 9’s.”

A numeric value is useful for communications between clinicians and provides a system that is well understood universally across the medical system. Numeric scales are especially useful in the context of postsurgery pain when clinicians are not as interested in the number itself as in whether the medication or treatment has been effective in reducing pain. In fact, the numeric scale was designed to provide a system for clinicians to note progression in acute and curable pain. Still, some clinician participants stated that in the emergency room, there is some aversion to the numeric system, because patients may exaggerate or falsify their pain score to receive treatment. There was a general belief from respondents that the emergency room sustains the problem of addiction because they cannot deny opioid treatment to patients who report high pain scores, especially if they have an outpatient opioid prescription.

**Phase II: Ideation (Generate Concepts and Designs)**

**Painimation Concept Development**

Based on the thematic insights taken from analysis of the user and stakeholder interviews, we established a set of design principles as criteria to support the creation of concept storyboards. A successful solution to the pain communication problem would meet all criteria listed in Textbox 2.

<table>
<thead>
<tr>
<th>Design principles as criteria.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aid patient in describing pain</strong></td>
</tr>
<tr>
<td>Given the scope of this project, the attempt was not to remove patient description or report altogether (with automated pain detection, for example) but rather to support that verbal description.</td>
</tr>
</tbody>
</table>

| **Quantitative representation of pain** |
|Patients want to know that their qualitative experiences matter as much as the quantitative selection. Clinicians, by contrast, required a number of some type that can indicate pain severity and show treatment-related improvements. |

| **Personalized** |
|Patients need to feel that assessment is personalized to them and their pain thresholds. With chronic pain it is all the more important to allow conversations to address the patient’s individual journey and take into account changes in their pain experience over time or even moment to moment. |

| **Concise** |
|Because time is limited (and pain assessment is just one part of the interaction between the patient and clinician), the procedure needs to be short and simple to complete, yet provide the necessary data to guide diagnosis and treatment. |

| **Facilitate the conversation** |
|Based on the study findings, the most prominent stakeholder need was for a tool that would improve the patient–provider interaction by making the communication surrounding pain symptoms easier, and helping patients feel heard and understood. The relationship between the patient and clinician was viewed as the most important aspect of the medical encounter. |

From the ideation session and concept selection process as described in the “Methods” section, 3 final concepts were selected and then developed out using generative storyboards to illustrate how the principles might fit into various scenarios. The 3 concepts were (1) expressive pain painimations, where patients would use animations to describe their pain to providers; (2) a personalized pain threshold scale, where rather than being restricted to a 0-10 scale patients would use an app to set their highest and lowest pain based on their own descriptors, words, or numbers; and (3) communication-style matching, where patients would be matched with a provider that fits their communication style. Needs validation sessions revealed that only the painimation concept qualified as both desirable and feasible for both patients and clinicians. Patients felt the painimations were more expressive than words or images alone, had an emotional quality, and even incorporated the fluctuations of pain over time. Clinicians and clinical researchers believed the painimation concept could work in their clinic and felt the concept would help create rapport between patients and clinicians.

In terms of feasibility, patients felt there might be individuals who prefer words over images or may not understand the painimations. It was evident that any tool would need to be very easy to understand. Clinicians and clinical researchers expressed that they would still need a number and method to translate the painimations into a score that can indicate severity or be used to compare with the traditional 0-10 numeric pain scale, or other pain assessment measures.

**Painimation Drawing Exercises**

As part of the painimation development process, a total of 16 graduate design students participated in a drawing exercise where they were asked to draw a series of pain adjectives from the McGill Pain Scale. The student drawings were then clustered based on approach and also arranged according to intensity. This exercise resulted in drawings that were quite similar. The drawings were grouped by similarity, and the final groupings were used to inform the initial set of painimations (Figure 3).
First, a low-medium intensity version was created for the initial 3 painimations (throbbing, cramping, and shooting). Next, 2 new painimations (pounding and tingling) were created, each with a high and medium value. These 2 words came from the original list and were created to provide more variety in the painimations to allow for a range of responses.

**Painimation Development**

In addition to the drawing exercise session, the graduate design students participated in a survey to evaluate a preliminary set of painimations based on early findings. The “what kinds of pain do you believe these animations evoke?” survey of throbbing, cramping, and shooting painimations revealed rich language within the responses, which were organized into emergent themes: recall, time + change, and representation (Textbox 3). Participants used the painimations as a starting point to recall pain incidents and memories. They mentioned the temporal or changing nature of pain. Additionally, participants indicated satisfaction and comfort using these painimations to represent a sensation.

The throbbing painimation had the highest responses of 1 particular word, which was “throbbing” (n=11). For the shooting painimation, “quick” and “sharp” had the same number of occurrences (n=5). The cramping painimation had a tie between “dull,” “deep,” and “slow” (n=2). Because of this lack of convergence, the cramping painimation was revised. Multimedia Appendix 1 displays word frequency in responses.
Textbox 3. Emergent themes from survey of painimations.

RECALL
Participants used the animations as a starting point to recall certain pain incidents and memories.
Example quotes:
Reminds me of when I was having my broken arm bent by a pair of nurses to be put into a cast.
Like when I come in from outside when it is cold and my ears heat up uncomfortably, or if I jam my finger and it swells to the point I can feel my heartbeat in my finger.

TIME + CHANGE
Participants mentioned the temporal or changing nature of pain.
Example quotes:
Pain that fluctuates in intensity.
Very erratic pulsing.
Something that starts out in one area and spreads across the body.
Coming up and then dying back down.
Slowly beginning with mild intensity, rising in a crescendo to a near-blinding, wince-inducing pain.

REPRESENTATION
Participants indicated satisfaction and comfort with using these animations to represent a sensation.
Example quotes:
This feels like it could describe that pain well.
I think the strong visuals might really speak to some people.
This could easily resemble how I felt when I got my wisdom teeth out.

Creation and Evaluation of Wireframes for a Painimation app
A new group of 5 patient participants and 4 clinician researchers was asked to interact with wireframes of an app that used painimations to measure pain. Feedback included participants wanting to see the whole set of painimations, so they knew how many choices they had. They also preferred that the intensity be depicted through a slider.

To resolve participant concerns, we created an instruction page to precede the viewing of actual images, on how to choose the painimations and increase and decrease the intensity; thumbnails of all painimations were shown on each screen with textual description, and arrows were replaced with a prominent slider. To provide users with feedback after making their selections, a panel was added at the bottom where the chosen painimations could be dragged and dropped (Figure 4).
Phase III: Implementation (Prototype, Test, and Iterate)

Final Set of Painimations

Based on user feedback throughout Phase II, a final set of 8 painimations were developed and then independently reviewed by 2 graduate design students outside of the investigative team (see example 2 in Multimedia Appendix 2).

The 2 graduate students were asked to label each painimation using a provided set of pain adjectives. Their labeling of the painimations approximated the intended representations, confirming that a broad set of pain types was depicted as unique feature sets, with no overlap between them.

These final 8 painimations (Figure 5) were then reviewed by the patients (n=5) and clinical researchers (n=4). Patient participants felt these painimations would aid in their pain communication, and several statements suggested that the painimations resonated. Participants would look through the set of painimations, choose 1 or 2, and make statements such as “This one really feels like my headache, exactly!” Other general comments about the idea itself included “These painimations feel like the aha moment for me. Hopefully, doctors will see it soon, too,” and “Just knowing that doctors are asking us this question with a tool that comes closer to what we’re feeling, shows that they are being empathetic and less dismissive.”

While patients acknowledged the benefits of seeing something more qualitative and contextual, they were also concerned about the limitations of the current system: “What is to stop me from getting frustrated with this system in the same way that I currently get frustrated with the number system [wanting to increase the value of the slider to more than what is possible]?”

The clinical researchers’ main concern was that the painimations needed validated numerical values of intensity. Although each painimation entry produced a numerical value of 0-100 on the slider and the painimation quality type (eg, “throbbing”), these values would need to have reliable and credible numerical correlations with the traditional numerical pain scale (eg, a particular painimation calibrated at a certain level would equal an 8/10 on the numerical pain scale).
Pilot Testing Painimation With Adults With Sickle Cell Disease

To confirm the acceptability and usability of the painimations, we tested a prototype of a Painimation app designed in Phase II with a use-case sample (adults with SCD-related pain). Six African American adults (age range 24-32 years, 67% [n=4] female) with SCD and self-reported chronic pain completed a pain entry using the prototype Painimation app and were asked to provide a verbal evaluation of all 8 painimations in a modified think-aloud protocol.

The adults with SCD reported that the Painimation app is more engaging, easier to use, has less entry burden, and leads to more of a conversation compared with other pain assessment forms they have used in the past.

In response to the question, “Would you find this animation applicable to your pain?” 6/6 patients with SCD responded “Yes” for electrifying; 5/6 for stabbing; 4/6 for burning; 3/6 for cramping; 2/6 for shooting; and 1/6 for throbbing, tingling, and pounding. Interestingly, 1 participant mentioned that the burning painimation looks like beginning stages of sickle cell crisis. Another patient felt that most painimations were not “severe” enough to represent her pain.

The types of pain seemed to differ between patients; however, many of the patients described their pain as stabbing and pulsating, and they consistently described some of their pain as continuous. In terms of pain tracking, 3/6 patients tracked their pain in their phone or journal, while 2 only documented pain crises, rather than daily pain. One patient said his pain did not change, so he did not feel the need to track it.

These patients echoed what patients with chronic pain in our earlier interviews reported regarding communicating with providers about pain. They liked when they felt like doctors listened and cared but were discouraged when they did not feel heard, when doctors seemed as if they did not have empathy, or did not understand their condition.

Discussion

Application of Painimation

Successful medical care depends on effective communication between patients and clinicians regarding the patients’ health symptoms and the most appropriate therapeutic path [35]. Providers are unable to deliver quality medical care when they lack the tools to appropriately assess or interpret patient symptoms that are critical to diagnosis and treatment. This is especially true for the assessment and treatment of pain.

Through a human-centered design approach, our study discovered that patients with pain frequently have negative interactions with providers characterized by misunderstandings, negative accusations, and intimidation. A major cause of this breakdown in the patient–provider interaction is the challenge in communicating pain and feeling understood. Patients, clinicians, and researchers in this study reported that the current pain assessment approaches used in the medical setting fail to accurately capture or communicate patients’ pain experience, have limited effectiveness for guiding diagnosis and treatment, and may exacerbate breakdowns in communications between patients and providers. Other studies have also reported that measures oversimplifying the pain experience may lead to patients’ personal legitimacy being undermined and result in clinicians inadvertently contributing to chronic pain stigmatization [36]. Given the importance of patients feeling respected and supported by their clinicians, it is imperative to improve patient–clinician communication regarding pain [37].

To address this gap, the current human-centered design study resulted in the development of a novel pain assessment approach that leverages digital animations. The use of pain animations or painimations showed promise with a use-case clinical sample of adults living with SCD-related chronic pain. Our prior published study found that patients’ selection of painimations were correlated with their scores on validated scales, and yielded some evidence that painimations may have better diagnostic potential than traditional multidimensional pain scales [17].
Given that pain is incredibly complex and its qualities are particularly difficult to express [8], there have been efforts to improve communication of pain [38-41]. For example, presenting abstract or literal pain images to patients with chronic pain during pain consultations was associated with clinician warmth and empathy, improving the patient–clinician rapport and communication [38]. Another example is Pain QuILT (a newer version of the Iconic Pain Assessment Tool), a web-based and mobile-accessible tool for the visual self-report and tracking of pain that offers 16 pain qualities, such as burning, electrical, and stabbing [39,40]. Pain QuILT was rated significantly easier to use than both the McGill Pain Questionnaire and the Brief Pain Inventory and was associated with fewer barriers to complete [40]. Our findings support and extend this work.

Abstract painimations can capture the experience of pain in a comprehensive manner. These painimations can be visually configured to reflect pain location, quality, and intensity. Moreover, they allow users to interpret the painimations instead of restricting them to specific/labeled pain quality options. The abstract and nonverbal nature of the painimations is also important because it helps level the playing field for marginalized or underserved populations. Patients with lower health literacy, communication disorders, or cultures/languages different from those of the providers have previously faced a communication gap that put them at a disadvantage when seeking medical care. While there are complicated power dynamics between a patient and a clinician, it benefits the patient to have a tool that does not rely on literacy or language, upon which to build conversation and allow patients to more effectively report their symptoms. As evidenced by this project, providing something that is removed from medical jargon or systems (which were not designed from a patient-centric perspective) allows patients to express themselves comfortably, knowing that their comments are valued, heard, and hopefully understood. Furthermore, these painimations address the disparities that current pain assessments perpetuate due to their use of complex words that may alienate individuals with low literacy, disabilities, cognitive impairment, or other communication barriers [12,13,42].

Relevance and Importance of Human-Centered Design Work

Human-centered design and evidence-based data, together, have significant potential for disease prevention and management [43]. Patients need to have the opportunity to participate as true partners in their health care [44]. Utilizing user-centered participatory approaches allows the evaluation of which elements work best for which populations in which contexts [45]. Thus, application of human-centered design in health care will exponentially improve the effectiveness of medical care and disease prevention [43].

Human-centered design is gaining traction in health care and the proliferation of mobile technologies expands opportunities for innovation, particularly because of the wide access to smartphones in clinical populations [23,46-48]. Mobile technologies have been shown to be beneficial in reducing pain severity and are well liked by patients and clinicians [49]. In fact, a study of perspectives of patients with chronic pain on methods of assessing pain found that 80% favored use of a digital version of body template/diagram, and 43% favored use of technology [50]. However, most mobile pain technologies (around 70%) still do not systematically engage patients with chronic pain as end users during app development, nor do they involve clinicians [51]. To ensure short- and long-term engagement of mobile app or digital health interventions, it is critical to include patients and clinicians in all stages, particularly the development stages [48,52-54].

Strengths and Limitations

This study has several strengths, including a rigorous human-centered design approach that involves target users and stakeholders at each phase. A major limitation of this study approach, however, is the small sample and thus limited age, genders, ethnicity/race, and number of pain conditions that were represented by the user and stakeholder groups. For example, only a small number of African Americans with SCD tested the app. Consequently, the generalizability of our findings is limited. The Painination concept will need to be tested by a larger, more representative sample in terms of age, gender, and ethnicity/race with a broad range of pain types, to determine if all pain experiences are represented in the current set of 8 painimations or if additional painimations need to be designed.

Finally, the reflexivity of the investigators and consultants needs to be considered and was systematically evaluated. It is likely that prior experiences and biases may have influenced the direction of designs and how the findings were interpreted. Future work in this area will benefit from more objective evaluations of the tool and the results.

Future Directions

This study demonstrates the process of human-centered design to build empathy for the end user and ultimately develop and implement an innovative solution for a prominent problem in medical care. Further research is needed to establish whether developing animations that explicitly measure affect and emotion would be beneficial. Additionally, how particular pain characteristics (conditions) might influence the further development of alternative methods (including this one) needs to be considered.

While these painimations have proven to have resonance with participants in this study, there is potential with augmented and virtual reality to develop the pain assessment experience further. For example, a doctoral project at the Norwegian University of Science and Technology in Trondheim is exploring how virtual reality can help nurses develop and sustain their empathy, as clinicians may become desensitized. It simulates morning sickness (nausea and dizziness, for example) through a headset that nurses wear. In relation to this project, the use of painimations in the virtual reality space, recreating the nausea and vomiting experience would be beneficial. Additionally, how particular pain characteristics (conditions) might influence the future development of alternative methods (including this one) needs to be considered.

Finally, with the current data, it is unclear whether Painimation is a tool to replace other measures or to be used in conjunction with other forms of pain assessment. Further, in clinical medicine the 0-10 scale is well-established as the status quo.
and health professionals will need to be convinced that using painimations offers useful and relevant information that can improve their clinical practice. Changing the pain assessment landscape is challenging and there are significant barriers to implementing new tools into routine clinical care. The current body of studies does not address how the pain conversation can be changed in this radical new direction; however, this is a starting point with potential to encourage and inspire other pain researchers to explore novel methods for assessing pain.

Conclusions
This study provides evidence that employing a human-centered design approach in clinical research has the potential to change how medical care is practiced. Currently, most electronic patient-reported outcomes measures for pain are essentially digital copies of paper–pencil questionnaires. Computer adaptive testing has helped streamline assessments, but the fundamental method of assessing symptoms and outcomes with words and numerical scales has not advanced along with the digital era. There is a need for more human-centered design studies to explore how technology can be leveraged to radically improve and advance how patient-reported pain outcomes are assessed.

Acknowledgments
The authors thank Andrea Ball, MA, for her professional medical writing assistance in the preparation of this manuscript.

Conflicts of Interest
CJ is exploring the potential to commercialize Painimation and may open a startup company based on Painimation in the future, but currently receives no revenue or proceeds from Painimation or any products related to this innovation.

Multimedia Appendix 1
Word frequency in responses from the Painimation survey.
[DOCX File, 77 KB - humanfactors_v9i1e27689_app1.docx]

Multimedia Appendix 2
Example of two painimations.
[MOV File, 7070 KB - humanfactors_v9i1e27689_app2.mov]

References


Abbreviations
CDC: Centers for Disease Control and Prevention
SCD: sickle cell disease

Please cite as:
PMID: 34994697

Edited by A Kashniruk; submitted 03.02.21; peer-reviewed by S Badawy, Y Virella Perez, B Thompson, R de la Vega; comments to author 03.03.21; revised version received 17.03.21; accepted 06.09.21; published 07.01.22.
Clinician Perspectives on Unmet Needs for Mobile Technology Among Hospitalists: Workflow Analysis Based on Semistructured Interviews

April Savoy1,2,3, PhD; Jason J Saleem4, PhD; Barry C Barker1, MA; Himalaya Patel1, PhD; Areeba Kara5,6, MS, MD

1Center for Health Information and Communication, Health Services Research and Development Service, Richard L Roudebush Veterans Affairs Medical Center, Indianapolis, IN, United States
2Purdue School of Engineering and Technology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States
3Center for Health Services Research, Regenstrief Institute, Inc, Indianapolis, IN, United States
4Department of Industrial Engineering, James Breckenridge Speed School of Engineering, University of Louisville, Louisville, KY, United States
5Indiana University Health Physicians, Indianapolis, IN, United States
6Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
April Savoy, PhD
Center for Health Information and Communication
Health Services Research and Development Service
Richard L Roudebush Veterans Affairs Medical Center
1481 West 10th Street, Stop 11H
Indianapolis, IN, 46202
United States
Phone: 1 3172782194
Email: asavoy@iu.edu

Abstract

Background: The hospitalist workday is cognitively demanding and dominated by activities away from patients’ bedsides. Although mobile technologies are offered as solutions, clinicians report lower expectations of mobile technology after actual use.

Objective: The purpose of this study is to better understand opportunities for integrating mobile technology and apps into hospitalists’ workflows. We aim to identify difficult tasks and contextual factors that introduce inefficiencies and characterize hospitalists’ perspectives on mobile technology and apps.

Methods: We conducted a workflow analysis based on semistructured interviews. At a Midwestern US medical center, we recruited physicians and nurse practitioners from hospitalist and inpatient teaching teams and internal medicine residents. Interviews focused on tasks perceived as frequent, redundant, and difficult. Additionally, participants were asked to describe opportunities for mobile technology interventions. We analyzed contributing factors, impacted workflows, and mobile app ideas.

Results: Over 3 months, we interviewed 12 hospitalists. Participants collectively identified chart reviews, orders, and documentation as the most frequent, redundant, and difficult tasks. Based on those tasks, the intake, discharge, and rounding workflows were characterized as difficult and inefficient. The difficulty was associated with a lack of access to electronic health records at the bedside. Contributing factors for inefficiencies were poor usability and inconsistent availability of health information technology combined with organizational policies. Participants thought mobile apps designed to improve team communications would be most beneficial. Based on our analysis, mobile apps focused on data entry and presentation supporting specific tasks should also be prioritized.

Conclusions: Based on our results, there are prioritized opportunities for mobile technology to decrease difficulty and increase the efficiency of hospitalists’ workflows. Mobile technology and task-specific mobile apps with enhanced usability could decrease overreliance on hospitalists’ memory and fragmentation of clinical tasks across locations. This study informs the design and implementation processes of future health information technologies to improve continuity in hospital-based medicine.

(JMIR Hum Factors 2022;9(1):e28783) doi:10.2196/28783
KEYWORDS
electronic health records; hospital medicine; user-computer interface; human-computer interaction; usability; mental workload; workflow analysis

Introduction

Electronic health record (EHR) systems aid documentation, information retrieval, and order creation. However, their lack of portability hampers effective support of communication between health care professionals and optimal access to patient information [1-3]. Such deficiencies contribute to task redundancies, constrain medical decisions at the point of care, and create inefficiencies that detract from valuable clinician-patient interactions [4-6]. These deficits are perhaps most impactful for hospitalists, a medical subspecialty focused on inpatient needs [7,8]. Multiple factors, including high patient acuity, ineffective health information technology (IT), hospital layouts, organizational policies, and interruptions, make hospitalists’ workflow cognitively demanding and dominated by activities away from the patient’s bedside (indirect care [9-15]).

As smartphones and tablets (mobile technology) became ubiquitous, they were proposed as one way to improve health IT. Physicians in emergency departments anticipated that these devices would improve workflow and physician-patient interactions [16], and in 2012, 87% of physicians were using smartphones and tablets in the workplace [17]. However, users in health care settings report lower expectations of mobile devices after actual use [17-19]. Most studies report experiences of physicians in training or those working in emergency departments. Less is known about the perceptions of hospitalists or their unique needs [20,21].

To improve care for Veteran patients, the US Department of Veterans Affairs (VA) Mobile Health Provider Program was launched in 2014. Through this program, over 12,000 iPads have been distributed at more than 60 VA sites. The program used a multiphase implementation strategy, focused on infrastructure updates, secure access to native mobile apps, and development of VA provider apps. However, adoption and use of the iPads and mobile apps among hospitalists has been low [22]. Our objective was to describe the needs and opportunities for mobile technology during the hospitalist workday. To characterize mobile technology that can synergistically support the workflow of hospitalists, we interviewed hospitalists to gain their perspectives on integrating mobile apps.

Methods

Overview

We conducted semistructured interviews guided by the Systems Engineering Initiative for Patient Safety (SEIPS) framework [5,23]. This framework, consisting of five factors (people, environment, tasks, tools, and organization) and their interactions, can be used to describe how health care providers’ work systems impact workflows and outcomes [5]. Our interviews focused on tasks from multiple workflows to obtain in-depth information about related frequencies, redundancies, difficulties, and mobile apps. Our analysis aimed to characterize contributing work system factors, multiple impacted workflows, and participants’ ideas for mobile app interventions.

Participants and Setting

The study was conducted at a 200-bed urban teaching hospital operated by VA in Indiana. This hospital offered iPad tablets and introductory training to its health care providers. We sought approximately 12 participants to increase the likelihood of thematic saturation [24,25]. We sought participants who practiced according to the hospitalist model of care because they may face overlapping workflow challenges. Physicians and nurse practitioners from hospitalist and inpatient teaching teams and second- or third-year internal medicine residents were eligible to participate [26,27]. Eligible participants were contacted via email, and a nonfiction book was offered for participation. This study was approved and overseen by the Institutional Review Board at Indiana University (#1608865326) and the Research and Development committee at Richard L. Roudebush VA Medical Center.

Semistructured Interviews

Data Collection

Semistructured interviews were designed as 45-minute sessions (Multimedia Appendix 1). Participants were asked to describe their primary roles and information-intensive tasks. They were then asked to identify tasks that were frequent, redundant, and difficult, and to explain their choices [28,29]. Definitions were reviewed with participants as follows:

- Information-intensive tasks: require reading, writing, or sharing information (eg, chart review)
- Frequent: performed often or for each patient (eg, looking up patients’ contact information or reviewing discharge summaries)
- Redundant: done repeatedly that should only be done once or not at all (eg, repetitious log-ins or clicks to access required information)
- Difficult: require uninterrupted time and attention (eg, reviewing labs or determining trends in vitals)

Participants were asked to describe the context of each task with a focus on work system factors [5]. Interviewers diagrammed discussions as participants spoke. Participants completed a demographic survey, including reporting use of self- and work-furnished mobile technology. Demographics and field notes were collected on paper and scanned. Interviews were audio recorded and transcribed.

Workflow Analysis

For each task identified as frequent, redundant, or difficult, we open-coded participants’ responses and organized these codes within the five work system factors. Next, we analyzed the impact those tasks had on workflows [30,31]. Lastly, we analyzed participants’ responses to mobile technology to identify and describe types of potential mobile app interventions.
Using a hybrid deductive-inductive approach, we iteratively developed a code book with sections and codes to aid each type of analysis [32]. We used a deductive approach to identify relevant work system factors and an inductive approach to describe workflow effects and potential mobile app interventions. One analyst created the preliminary code book based on the SEIPS work system factors and open coding of three transcripts. Using this preliminary code book, four additional analysts reviewed another set of three transcripts. The team discussed and refined codes. With the revised code book, four analysts worked in pairs to code the remaining transcripts, which were randomly assigned. After coding each transcript independently, coding partners reviewed transcripts line by line, resolving discrepancies through consensus meetings. If new codes emerged during coding, they were retroactively applied to previously coded transcripts. Each analyst wrote memos for prominent codes; then, analyst pairs conducted consensus meetings. After these meetings, an analyst selected the most frequent open codes, linked the most frequently co-occurring codes for each, and prepared a narrative summary with supporting quotations. Coding, memo writing, and content analysis were performed using Excel (Version 2016, Microsoft Corporation).

Results

Participants

Over 3 months, we interviewed 12 participants: 9 staff physicians, 1 resident physician, and 2 nurse practitioners. Including residency, experience ranged from 0.9 to 37 years (mean 11.7, median 8.5); experience in the present organization was similar (0.5-37 years; mean 10.8, median 8.5). A total of 11 (92%) reported using mobile technology at work, including both personally owned and work-furnished devices. Only 3 (25%) reported using iPad tablets at work. Nurse practitioners worked on the hospitalist team, while staff physicians rotated between the nonteaching hospitalist team and inpatient teaching teams. Patient load was estimated as ranging from an average of 10 to 15 patients per day.

In the following section, we first present participants’ perceptions of specific tasks that were perceived as frequent, redundant, and difficult. Next, we describe the workflows that were perceived to be most impacted by these tasks. Finally, we present participants’ perceptions of mobile technology and potential mobile app interventions.

Frequent, Redundant, and Difficult Tasks

Chart reviews, orders, and documentation were identified as the most frequent, redundant, and difficult tasks.

Chart Reviews

Participants described that chart reviews (re)established the patient’s trajectory, which included viewing patient history, recent notes, laboratory results, and vitals. Participants reported conducting a summary review for every patient throughout the day to monitor progress, orders, procedures, and test results, and estimated spending 30 minutes to review the chart of a new patient. Some participants noted that their initial reviews were completed at the beginning of the day or before their shift. This need for on-demand continuity was contrasted with fragmentation of records in EHRs and the multistep methods for accessing them. Participants characterized chart review as redundant because of intermittent updates without notification, resulting in checking either too little or too often. Information copied in workrooms and carried to patients could be outdated upon reaching patients, or effort could be wasted looking for information that had not yet arrived.

Orders

Participants described writing orders multiple times a day using computers. Orders included lab tests, consultations, and prescriptions. The institution currently requires electronic entry of all orders. Participants described the lack of (bedside) computers, not necessarily the need for complex thought, was what made ordering difficult and inefficient. Perception of ordering was also negatively affected by poor EHR usability. The organization of orders in the EHR was thought to be unclear, decreasing the discoverability of specific orders. Participants gave examples of order forms for similar procedures that were found in different branches of the menu. This poor organization of order forms was described as increasing difficulty by limiting application of knowledge between orders—finding and writing one type of order did not necessarily make it easier to find or write other types.

Documentation

Documentation was reported as one of the most labor- and time-intensive tasks. It included documenting a variety of information, including histories and physicals, visit notes, daily note, and discharge summaries. As with chart review, fragmentation of information in EHRs meant writing notes frequently, even when new notes were similar to previous notes. Participants described EHR documentation as a constant process consuming a considerable portion of the day. Patient load was estimated as ranging from an average of 10 to 15 patients per day. With that, participants estimated that documentation time averaged 30 to 45 minutes per patient. Participants described documentation as redundant, as they and their trainees were required to write notes for the same patients.

Table 1 summarizes the contributing work system factors for each task perceived as frequent, redundant, and difficult with some illustrative quotes from participants.
Table 1. Frequent, redundant, and difficult tasks identified by participants and derived contributing work system factors.

<table>
<thead>
<tr>
<th>Task</th>
<th>Contributing work system factor(s)</th>
<th>Representative quotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chart reviewa</td>
<td>• People: Extent of reliance on electronic records varied between participants</td>
<td>“Ideally, you would like to be able to harvest that information in the room with the patient by handheld device so that if memory fails and patients have questions, you can use that to help answer their questions. Mostly, I do that from memory now.”</td>
</tr>
<tr>
<td></td>
<td>• Environment: Electronic chart was not accessible at bedsides</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tools/technology: EHR did not push notifications of important changes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tasks: Patients with more status changes needed more frequent review</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Organization: Multifactor authentication was required before every EHR session</td>
<td></td>
</tr>
<tr>
<td>Ordersb</td>
<td>• People: Preferences varied in when to start and when to submit orders</td>
<td>“There are multiple clicks to get to different boxes, lots of pop-ups that you have to go through...the computer system itself adds considerably to the amount of time that we take and takes away from our patient care”</td>
</tr>
<tr>
<td></td>
<td>• Environment: Electronic ordering was not accessible at bedsides</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tools/technology: Finding the right order form in the EHR was difficult</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tasks: Orders depended on having the most up-to-date patient information</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Organization: All orders had to be made through the EHR</td>
<td></td>
</tr>
<tr>
<td>Documentationc</td>
<td>• People: Content of attendings’ notes depended on the content of their residents’ notes</td>
<td>“I think documentation is by far the thing that takes us the longest— documentation for sure.”</td>
</tr>
<tr>
<td></td>
<td>• Environment: EHR was not accessible at bedsides</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tools/technology: Authoring notes in the EHR sometimes involved copying forward text from older notes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tasks: These were sometimes based on a single encounter, and other times more longitudinal (eg, discharge summaries)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Organization: Facility required a series of documentation and ordering steps before discharge</td>
<td></td>
</tr>
</tbody>
</table>

aChart review: going through patient information and history.
bEHR: electronic health record.
cOrders: services like lab tests and referral.
dDocumentation: summarizing encounters, making or changing care plans, and adding to patient information.

Impact of Tasks: Inefficient and Difficult Hospitalists’ Workflows

Participants characterized admit, discharge, and rounding workflows as difficult or inefficient.

Intake

The admit workflow was reported to be time-consuming:

*It takes 1.5-2 hours to do an admission from start to finish,...entails chart review, seeing the patient, putting in orders, reviewing things, and doing the history and physical.*

Difficulty of completing tasks seemingly increased as the workday progressed. Often, patients’ care was distributed across multiple health care systems. In those cases, admitting was described as involving retrieving both internal and external records. At best, external records were retrieved electronically (eg, from a health information exchange). Otherwise, retrieving outside records involved making telephone calls and reviewing scanned records. Some participants relied on residents:

*I usually have learners helping take care of some other tasks but without learners sometimes it [compiling patients’ histories] just doesn’t happen.*

Discharge

Most participants noted efforts to complete discharges by lunchtime:

...it’s usually like a 4-5 hour process. It’s challenging to discharge patients in the afternoon, because there’s just too much to get done. It’s cumbersome...

Due to documentation demands, participants described these workflows as redundant and time-consuming. Discharges involved data retrieval that depended on the length of the stay and much documentation. Several notes needed to be written, and among those notes, a large amount of information was duplicated:

...So discharge note, anticipated note, discharge instruction, discharge summary, medical reconciliation, pharmacy output...we can clump together to save time...

These characteristics related to admit and discharge workflows increased participants’ time in workrooms because access to their desktops were required to complete notes.

Rounds

Rounding was identified as an inefficient workflow. Participants reported seeing 10 to 18 patients during rounding. For each patient, participants documented history and physical notes in
the EHR. Afterward, they duplicated the text on paper cards to support rounding. Otherwise, the information was not readily available. Printouts were wasteful because page counts averaged 6 pages per patient. Unlike computers and the EHR, paper notes fit in their pockets and were easily accessible for review and modification:

Bedside computers are not at the bedside. We don’t really have access to computers that work very well other than those in our team rooms.

Based on participant interviews, Figure 1 illustrates a snapshot of the participants’ description of difficult and inefficient workflows that stem from frequent, redundant, and difficult tasks.

Figure 1. Snapshot of difficult and inefficient elements in hospitalists’ workflows. Hospitalists start in the charting room and conduct chart reviews for all patients they will visit. For each patient, hospitalists must duplicate information from the EHR on index cards or printouts to support review at bedside. After completing cards, hospitalists take all the cards to the ward where patients are located. Hospitalists find the appropriate card for each patient encounter and update the card with new patient information related to status, orders, and plans. Hospitalists move from one patient to the next, repeating those steps. After the last patient encounter, hospitalists go back to the charting room to enter the information from the cards into the EHR. This entire workflow is done multiple times a day. EHR: electronic health record.

Participants’ Perspectives on Opportunities for Mobile App Interventions
Participants expected mobile technology to decrease task completion time; however, they noted that neither rapid access nor documentation of information was supported by current mobile apps. Usability issues were also noted, highlighting the misalignment between expected and actual functionality. One participant said:

I had an iPad for a while here when I was in the pain clinic, but I didn’t use it. I couldn’t do controlled substances refills on it, and that was all that I ordered in the pain clinic. So I turned it back in...

As a result, participants expressed that they viewed paper and their brains as “surrogate mobile devices.” Paper to-do lists were described as repositories of patient information and task trackers. Paper was perceived as more reliable than mobile technology. Alternatively, some participants described heavy reliance on their memory. As one participant noted, pointing to his head, “my technology is up here.”

When discussing ideas for mobile technology, participants prioritized portability, reduction in task time completion and task completion at bedside (Textbox 1). Three representative examples of useful mobile apps emerged. First, participants said a mobile device like an iPad would help patient-provider communication and entering orders at bedside. Second, participants described a note-taking app that had sharing features and stored nurses contact information. Lastly, participants proposed an app for electronic consents.
Textbox 1. Participant quotes describing potential mobile app solutions.

<table>
<thead>
<tr>
<th>Patient-provider communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>“…instead of telling patients, actually giving them a visual as you are rounding will make them feel more involved in their own care.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>“…[communication] breakdown occurs when we’re calling nurses…if we just had the correct number in the first place, we wouldn’t have to go through talking to multiple people.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>eConsent</th>
</tr>
</thead>
<tbody>
<tr>
<td>“…when you go buy a coffee and doughnuts, you know how you can just sign on the iPad; having that same setup for consent may work well.”</td>
</tr>
</tbody>
</table>

Discussion

Summary

Provider-EHR interactions in inpatient care have contributed to increased workflow inefficiencies, reduced time for provider-patient interactions, and increased cognitive burden among hospitalists. Our findings provide a better understanding of the misalignment between hospitalists’ needs and expectations of available mobile technology and evidence of hospitalists’ cognitively intense workflows. In this section, we discuss our findings and implications for implementation of future mobile technology interventions for hospitalists.

Influential Contextual Factors

Despite the need for mobile access to patient information, mobile technology was not widely adopted. Difficult and inefficient tasks were predominately related to provider-EHR interactions because access to EHRs was not consistent at bedside [33]. We associated hospitalists’ unmet needs with one or more of the following SEIPS factors: tasks, tools and technology, and environment (location). For example, participants had to travel across three floors to complete workflows and clinical tasks that required chart review, patient encounters, and documentation. These dynamics influenced the perceived difficulty, frequency, and redundancy among workflows and clinical tasks. Although mobile technology was available, usability issues related to existing mobile apps prevented their use, increasing participants’ reliance on index cards and printouts. Thus, contextual factors influenced the need for mobile technology, but misalignments of hospitalists’ expectations and mobile device functionality limited the adoption and use of existing mobile apps. This finding demonstrates the critical importance of integrating workflow analysis into the design process of mobile technology interventions; the result of this analysis identifies unmet needs and unintended consequences.

Implications of Cognitive Workload and Burden

The three information-intensive tasks (chart review, ordering, and documentation) identified as frequent, redundant, and difficult were prone to an overreliance on hospitalists’ memory, including working memory [34]. A major contributor to this overreliance was the lack of mobile technology that supported chart review or order entry needed at bedside. Classifying tasks as frequent and redundant were easy for participants. These tasks were often described as inefficient and sources of hospitalists’ frustration. Identifying tasks as difficult caused participants to think of their tasks in a new way. Echoed throughout our data collection, tasks were not difficult due to hospitalists’ lack of knowledge or training to identify treatment plans, make clinical decisions, or perform clinical procedures. Rather, difficulty was defined and associated with the lack of support and access to usable technology required to review and enter information at bedside. According to existing cognition literature [35-37], these workflow aspects required participants to change location frequently, which increased the likelihood of interruptions and limited information recall (ie, cognitive slips and mistakes). This can be linked to incomplete documentation, communication breakdowns, and delays in care alluded to in participant interviews. Paper-based work-arounds were associated with processing orders and notes together in one sitting (ie, batch processing), not individually at the time of each decision. Batch processing has been associated with delayed team communication, delayed discharges, and shift limit violations [38]. Although the terms cognitive burden and mental overload were not specifically mentioned in interviews, these were clear outcomes for hospitalists based on our analysis. Cognitive burden can decrease resilience, situation awareness, and subsequently patient safety [36,39-41].

Potential for Task-Specific Mobile Apps

Hospitalists thought task-specific apps would be most helpful. Their primary goals were to reduce inefficiencies or difficulties with orders, discharge, consent, and team communication. Hospitalists’ focus on individual tasks indicates a need to shift design goals of mobile apps that focused on granting access to the entire EHR via consistent user interfaces (eg, mobile version of EHR desktop interfaces). Participants stressed the need for task-specific apps that highlighted fast, focused technology interactions when away from the charting room. The design of mobile apps should be based on the objective and use of the paper or cards currently used for hospitalists’ mobile workflows, including quick review and documentation of prioritized patient information. For example, apps should present customized views of each decision. Batch processing has been associated with processing orders and notes together in one sitting (ie, batch processing), not individually at the time of each decision. Batch processing has been associated with delayed team communication, delayed discharges, and shift limit violations [38]. Although the terms cognitive burden and mental overload were not specifically mentioned in interviews, these were clear outcomes for hospitalists based on our analysis. Cognitive burden can decrease resilience, situation awareness, and subsequently patient safety [36,39-41].

Potential for Task-Specific Mobile Apps

Hospitalists thought task-specific apps would be most helpful. Their primary goals were to reduce inefficiencies or difficulties with orders, discharge, consent, and team communication. Hospitalists’ focus on individual tasks indicates a need to shift design goals of mobile apps that focused on granting access to the entire EHR via consistent user interfaces (eg, mobile version of EHR desktop interfaces). Participants stressed the need for task-specific apps that highlighted fast, focused technology interactions when away from the charting room. The design of mobile apps should be based on the objective and use of the paper or cards currently used for hospitalists’ mobile workflows, including quick review and documentation of prioritized patient information. For example, apps should present customized views of patient health status or trends. In addition, apps can support bedside order entry with smart templates that use automation or dictation to optimize data entry without keyboards. Based on our findings, mobile apps designed to support the iterative nature of hospitalists’ workflows or rounds by providing a means to review charts and document at bedside may reduce the need for batch processing before and after patient encounters. Thus, current workflows would be streamlined, decreasing the redundancy and difficulty illustrated in workflows characterized in our study.

There are several task-specific apps that are being trialed and should be monitored for success. Since the completion of our
study, the VA’s Office of Connected Care is working to achieve greater understanding of provider preferences for mobile technology and task-specific apps. Providers currently have access to a variety of task-specific apps for mobile computing through the VA App Store. For example, the Image Viewing Solution is an app to access diagnostic-grade images. Annie App for Clinicians allows providers to assign disease-specific protocols to their patients. Several other task-specific apps are in development to meet VA providers’ needs.

Limitations

This workflow analysis was limited by a relatively small sample in one health care facility. VA is the nation’s largest integrated health care system. Therefore, participant perspectives of hospitalists’ workflows and mobile technology may be broadly relevant to other health care systems. For example, initial deployment of mobile technology, without a variety of task-specific clinical apps readily available contributed to the low adoption of mobile technology [18,19]. By using informal definitions of frequent, redundant, and difficult, these concepts may have overlapped to some degree. We did not explore differences associated with career stage (eg, early, middle, and late). Our findings demonstrated the influence of contextual factors; future studies should further explore interactions between technology use, interruptions, and geographic cohorting across multiple facilities [35,42-44].

Conclusion

Based on our results, there are opportunities for mobile technology to decrease the difficulty and increase the efficiency of hospitalists’ workflows. Mobile technology and task-specific mobile apps with enhanced usability have the potential to decrease overreliance on hospitalists’ memory and fragmentation of clinical tasks across locations that exist with current health IT and hospital environments. Task-specific apps that aim to reduce redundancies or excessive administrative work related to admissions, orders, and discharges were prioritized by hospitalists. Human factors engineering approaches are needed to identify hospitalists’ requirements for mobile technology to address issues with information management and recall during rounds. Extending beyond hardware features, a better understanding of direct and contextual factors of mobile information needs is required to develop mobile apps that can support hospitalists’ workflows. This will be influential in initial and sustained adoption of future mobile technology and apps.

Acknowledgments

This work was supported by a pilot grant (PPO 15-401; AS) and a Center of Innovation grant (CIN 13-416, M Weiner), both from the United States Department of Veterans Affairs Health Services Research and Development. AS is supported in part by the following grants: KL2TR002530 (A Carroll, PI), and UL1TR002529 (A Shekhar, PI) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award.

Authors’ Contributions

AS proposed the study and secured funding. BCB wrote the interview guide with input from AS, HP, and research assistants (Diana Natividad and Rachel Dismore). BCB led interviews and analysis. AS drafted the manuscript. Anna Mathew created the workflow illustration. All authors interpreted the findings, made critical revisions, and approved the published manuscript; all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.

[DOCX File, 16 KB - humanfactors_v9i1e28783_app1.docx ]

References


Abbreviations

- **EHR**: electronic health record
- **IT**: information technology
- **SEIPS**: Systems Engineering Initiative for Patient Safety
- **VA**: Veterans Affairs
Key Challenges and Opportunities for Cloud Technology in Health Care: Semistructured Interview Study

Kathrin Cresswell¹, PhD; Andrés Domínguez Hernández², PhD; Robin Williams³, PhD; Aziz Sheikh¹, MD

¹Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom
²Department of Computer Science, University of Bristol, Bristol, United Kingdom
³Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Kathrin Cresswell, PhD
Usher Institute
The University of Edinburgh
Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 131 651 4151
Email: Kathrin.Cresswell@ed.ac.uk

Abstract

Background: The use of cloud computing (involving storage and processing of data on the internet) in health care has increasingly been highlighted as having great potential in facilitating data-driven innovations. Although some provider organizations are reaping the benefits of using cloud providers to store and process their data, others are lagging behind.

Objective: We aim to explore the existing challenges and barriers to the use of cloud computing in health care settings and investigate how perceived risks can be addressed.

Methods: We conducted a qualitative case study of cloud computing in health care settings, interviewing a range of individuals with perspectives on supply, implementation, adoption, and integration of cloud technology. Data were collected through a series of in-depth semistructured interviews exploring current applications, implementation approaches, challenges encountered, and visions for the future. The interviews were transcribed and thematically analyzed using NVivo 12 (QSR International). We coded the data based on a sociotechnical coding framework developed in related work.

Results: We interviewed 23 individuals between September 2020 and November 2020, including professionals working across major cloud providers, health care provider organizations, innovators, small and medium-sized software vendors, and academic institutions. The participants were united by a common vision of a cloud-enabled ecosystem of applications and by drivers surrounding data-driven innovation. The identified barriers to progress included the cost of data migration and skill gaps to implement cloud technologies within provider organizations, the cultural shift required to move to externally hosted services, a lack of user pull as many benefits were not visible to those providing frontline care, and a lack of interoperability standards and central regulations.

Conclusions: Implementations need to be viewed as a digitally enabled transformation of services, driven by skill development, organizational change management, and user engagement, to facilitate the implementation and exploitation of cloud-based infrastructures and to maximize returns on investment.

(JMIR Hum Factors 2022;9(1):e31246) doi:10.2196/31246

KEYWORDS
cloud technology; qualitative; adoption; implementation; digital health; data processing; health care; risk assessment; user engagement
**Introduction**

**Background**

There is now an international drive toward digitally enabled, data-driven transformation of health care services, with health systems seeking to optimize work processes; improve the quality, safety, and efficiency of care; and reduce costs [1,2]. Health care typically relies on a web of complex information infrastructures that lack integration and interoperability, which contributes to fragmented service provision [3]. Such infrastructures may range from systems allowing data analysis within individual organizations to advanced cloud-based systems facilitating cross-organizational data-driven analysis [4].

Although the origins of cloud technology can be traced back to the 1960s, the term *cloud computing* has only emerged in this millennium [5]. It essentially involves delegating storage and processing of data to third-party organizations accessed via the internet rather than hosting them on an organization’s own computers. In doing so, cloud-based technologies can provide access to sophisticated large-scale technological infrastructures and advanced analytics services with the scope to rapidly scale up to meet peaks of demand [6]. Cloud product types differ in the degree of vendor and organizational control and can be public (shared across organizations), private (shared within organizations), or hybrid (a combination of both where on-premise infrastructure is combined with a public cloud). Hybrid clouds are increasingly popular as they not only allow access to public cloud infrastructure capacity but also maximize the use of on-premise solutions and therefore are a middle ground option for organizations with significant installed information technology (IT) capacity [7].

Textbox 1 summarizes the most common cloud products used in health care settings.

**Textbox 1.** Most common cloud products used in health care settings.

<table>
<thead>
<tr>
<th>Common cloud products</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Software as a service, where a cloud provider hosts software services that user organizations can access on the web (eg, a cloud-based electronic health record such as Athenahealth)</td>
</tr>
<tr>
<td>• Platform as a service, where providers make development tools available to the user via the cloud (eg, Microsoft Azure)</td>
</tr>
<tr>
<td>• Infrastructure as a service, where the service provider supplies cloud-based infrastructure components to the client, such as storage, servers, and networks (eg, Virtustream Enterprise Cloud)</td>
</tr>
</tbody>
</table>

Although cloud computing has transformed many industries (eg, entertainment and financial services) [8], its use in health care remains limited. There are some exceptions of promising developments in advanced health care systems that are now reaping the benefits (Textbox 2) [9]. The advantages of the cloud have been particularly visible in the wake of the COVID-19 pandemic, which has called for rapid deployment and cross-organizational integration of services as well as large-scale real-time data analytics [10].

**Textbox 2.** Examples of advanced health care systems that have implemented cloud technology.

<table>
<thead>
<tr>
<th>Examples of advanced health care systems that have implemented cloud technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The Shulan Health Management Group (China) implemented Amazon Web Services to host their “homegrown” system [11].</td>
</tr>
<tr>
<td>• The University of California, Los Angeles Health (United States) implemented Microsoft Azure for data processing and for integrating electronic health record data and data from other sources [12].</td>
</tr>
<tr>
<td>• The Mayo Clinic (United States) announced a strategic partnership with Google Cloud in 2019 [13].</td>
</tr>
</tbody>
</table>

However, despite some international governmental efforts to promote *cloud first* policies that foster the use of public cloud offerings in technology procurement [14,15], there are still significant points of friction in the adoption of cloud-based services. Some of these include concerns about security; fears of potential legal disputes between service providers and organizations; and issues surrounding vendor lock-in, privacy, ethics, and data ownership [16-20].

**Objectives**

In this study, we seek to understand how current opportunities in data-driven innovation facilitated by cloud computing could be positively harnessed in health care settings while minimizing perceived or actual risks.

**Methods**

**Overview**

We conducted a qualitative study between September 2020 and November 2020 using semistructured interviews sampling cloud providers, system implementers, software vendors, customers, and health informatics academics to gain an in-depth understanding of the evolving cloud ecosystem. It is important to keep in mind that data collection took place in the midst of the global COVID-19 pandemic and in the context of ongoing deliberations on the potential uses of cloud technology to address emerging urgent pandemic-related challenges. Discussions were strongly influenced by this topic.
Ethical Approval
We obtained ethical approval before the start of the study from the Usher Institute Research Ethics Group at the University of Edinburgh. Participants were provided with a consent form and an information sheet describing the study aims, procedures, and data management practices before participating in the study. They were given at least 48 hours to consider whether they agreed to participate and provided written informed consent. We informed the participants that they were free to withdraw at any time and that their responses would be anonymized during the analysis, removing names and places that could lead to identification of individuals.

Recruitment of Participants
We purposefully sampled stakeholders with perspectives on the topic of implementation, adoption, and optimization of cloud technology in health care settings [21]. Our aim was to gain a broad overview of different perspectives to understand the challenges and opportunities around cloud technology in health care settings and draw lessons that could inform future strategies for decision makers. In doing so, we specifically targeted individuals working across technology implementation, operations, design, research, and innovation within a range of organizations. We identified and recruited participants through our existing networks and communication channels as well as Google and LinkedIn searches using keywords related to the cloud and eHealth (eg, digital health, digital transformation of health, cloud computing, and cloud first). We complemented this strategy through snowball sampling by asking participants for recommendations of further interviewees. We aimed for variability in terms of geographical location (not including low- and middle-income countries as existing information infrastructures and challenges in these countries are likely to vary significantly), organizational function, area of expertise, and gender. Participants were selected based on their relationship with cloud technology in health care, both from the supply (cloud and software vendors) and demand (health care providers) sides. This included those who had experiences and opinions on the topic through experience of developing cloud solutions and cloud-enabled software, implementing and operating systems, or researching cloud technology.

Data Collection
ADH, a researcher with a background in science and technology studies and theoretical foundations surrounding information infrastructures, conducted all interviews via videoconference call software (Microsoft Teams). Interviews took the format of a conversation with a purpose where participants were encouraged to discuss issues important to them. ADH and KC (a social scientist with a background in sociotechnical theory) met periodically throughout the data collection process to discuss emerging findings and modify key lines of inquiry.

The interviews ranged in duration from 40 to 70 minutes. There were 20 one-to-one interviews and 1 group interview with 3 participants. Where participants asked for a group interview, we accommodated this request as it was more convenient for the participants and allowed us to gain insights into their complementary perspectives simultaneously. Although questions were tailored to individual roles and modified in line with emerging findings, we followed a topic guide exploring the state of cloud-enabled digital transformation in health care; views on barriers to realizing the potential benefits, risks, and areas of concern; and suggestions on how to address them (Textbox 3). During this process, the interviewer incorporated emerging themes across various interviewees and explored the tensions and differences in viewpoints in detail. We stopped collecting data when no new themes emerged during the concurrent analysis [22].

Textbox 3. Topic guide.

<table>
<thead>
<tr>
<th>Topic guide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewee’s background, current position, and description of the organization</td>
</tr>
<tr>
<td>Overview of the cloud ecosystem, stakeholders, and existing offerings</td>
</tr>
<tr>
<td>Implications of cloud ecosystem, stakeholders, and existing offerings</td>
</tr>
<tr>
<td>Promising and concrete use cases of cloud technology in health care</td>
</tr>
<tr>
<td>Challenges, risks, and hindrances for innovation in the cloud</td>
</tr>
<tr>
<td>Distinctive challenges of health care compared with other industries and sectors</td>
</tr>
<tr>
<td>Concerns about privacy, security, data ownership, and ethics</td>
</tr>
<tr>
<td>State of affairs and challenges in terms of integration and interoperability between cloud platforms</td>
</tr>
<tr>
<td>Role of the government</td>
</tr>
<tr>
<td>Future outlook (5-10 years) of the cloud in health care</td>
</tr>
</tbody>
</table>

Data Analysis
The interviews were transcribed using an external professional service and subjected to thematic analysis [23]. ADH verified the interview transcripts by listening to the audio recordings and correcting any inaccuracies before analysis.

We used a mixture of deductive and inductive thematic coding [24]. We added the transcripts to an NVivo (QSR International) version 12 project and theme coded them using a sociotechnical coding framework developed by the research team [25]. This framework highlights how different technological and social dimensions interrelate and how different perspectives shape...
aspects of the implementation and adoption of new technologies (Textbox 4). In addition, identified themes that did not fit the analytical framework were included in new categories.


<table>
<thead>
<tr>
<th>Dimensions used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology (the technological properties of the system and the surrounding infrastructure)</td>
</tr>
<tr>
<td>People (how various stakeholders use technology, including their expectations and experiences)</td>
</tr>
<tr>
<td>Organizations (how organizations implement technology and how this shapes use)</td>
</tr>
<tr>
<td>Macroenvironmental factors (how political and economic factors and markets shape technology development, use, implementation, and optimization)</td>
</tr>
</tbody>
</table>

ADH performed the first round of coding, periodically discussing emerging findings with KC. KC then re-examined the codes, resulting in minor changes to node titles and summarized the results in a narrative format. As part of our reflexive process, we identified how our previous experiences, assumptions, and preconceptions bore on the interpretation and coding of the data. In doing so, we discussed emerging findings within the research team to identify the relevance of themes within the Technology, People, Organizations, and Macroenvironmental factors (TPOM) framework as well as the need for new categories. We focused on examining converging and diverging perspectives, the interplay of technological and social dimensions, and the tensions and trade-offs emerging in the progress of cloud technology implementation, adoption, and optimization in health care settings.

Results

Overview

We interviewed 23 individuals (Table 1), including professionals working across major cloud providers, health care provider organizations, innovators, small and medium-sized software vendors, and academic institutions.

Table 1. Characteristics of the participants.

<table>
<thead>
<tr>
<th>Participant number</th>
<th>Gender</th>
<th>Location</th>
<th>Occupation</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>United States</td>
<td>Executive</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Software vendor</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Health care provider</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Software vendor</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Health care provider</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Software vendor</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Health care provider</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>France</td>
<td>Operations</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>11</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Academic</td>
<td>Health care provider</td>
</tr>
<tr>
<td>12</td>
<td>Female</td>
<td>United States</td>
<td>Operations</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>13</td>
<td>Female</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>14</td>
<td>Female</td>
<td>Finland</td>
<td>Academic</td>
<td>Research</td>
</tr>
<tr>
<td>15</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Software vendor</td>
</tr>
<tr>
<td>16</td>
<td>Female</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Research</td>
</tr>
<tr>
<td>17</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Health care provider</td>
</tr>
<tr>
<td>18</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Software vendor</td>
</tr>
<tr>
<td>19</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Software vendor</td>
</tr>
<tr>
<td>20</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Executive</td>
<td>Software vendor</td>
</tr>
<tr>
<td>21</td>
<td>Male</td>
<td>United Kingdom</td>
<td>Implementer</td>
<td>Health care provider</td>
</tr>
<tr>
<td>22</td>
<td>Female</td>
<td>United Kingdom</td>
<td>Operations</td>
<td>Cloud vendor</td>
</tr>
<tr>
<td>23</td>
<td>Male</td>
<td>United States</td>
<td>Executive</td>
<td>Cloud vendor</td>
</tr>
</tbody>
</table>
We produced 40 codes within the following four thematic areas: organizational context, social-human factors, technological factors, and wider macroenvironmental factors. The researchers then discussed the codebook and identified 4 salient challenges that were common across different interviewee backgrounds and affiliations. These were (1) drivers and perceived benefits associated with cloud technology in health care; (2) organizational and technological barriers limiting cost-effective use of cloud functionality; (3) infrastructural changes not immediately visible to frontline users, resulting in lack of clinical pull; and (4) visions of the future cloud vendor ecosystem.

Figure 1 illustrates how these emerging themes map onto the TPOM framework. As illustrated, the new emerging overarching categories related to cross-cutting issues spanning more than one TPOM dimension.

Drivers and Perceived Benefits of Cloud Technology in Health Care

The participants described various uses of cloud functionality, including scheduling software, videoconferencing, call center management, imaging analysis, and patient data analytics. On the basis of the most frequently discussed uses, we identified three salient categories: (1) organizational dimensions (eg, remote and collaborative working at scale, modeling algorithms surrounding predictive analytics, organizational analytics, and automation), (2) patient- and clinician-facing (eg, remote working, chatbots, and community outreach functionality), and (3) cross-organizational and regional (eg, data analytics surrounding particular disease areas for population health management and research).

We observed overall positive attitudes among the participants in relation to how cloud computing helped harness the value of data-driven innovation at scale. The adoption of cloud technology was perceived to be driven by existing issues faced by the sector, particularly concerning limited resources, access to and delivery of care, administrative workloads, and availability of critical services. Positive attitudes were particularly salient among system implementers, who saw immediate gains through secondary uses of data and tackling some of the most pressing challenges for health care posed by the COVID-19 pandemic. Here, cloud technology facilitated the deployment of solutions at speed without the need to purchase additional hardware:

Cloud enabled responsiveness...and throughout COVID, that’s what that’s been about. And it’s removed one challenge off getting hold of hardware, getting it set up and all the rest. So, it’s made us more responsive, it’s made us quicker to adapt...the forcing function was COVID, and cloud’s helped us have a faster response. [Participant 2, male, software vendor, United Kingdom]

Other key benefits associated with cloud technology mentioned by the participants included cost-effective management and storage of data at scale combined with ready access to advanced computing capabilities and tools, such as machine learning (ML) and natural language processing:

And one of the greatest things about this now is machine learning and AI [artificial intelligence]...it hasn’t been up until recently when [vendor] fully put a heavy effort over the last five or six years about democratizing access to these tools at scale, because you’re not only interested in building one or two models, you’re interested in building hundreds, thousands, tens of thousands of these models. [Participant 23, male, cloud vendor, United States]

Organizational and Technological Barriers Limiting Cost-effective Use of Cloud Functionality

Although benefits associated with cloud technology were realized in many organizations that the participants worked with, they also discussed how these might not be representative of the wider health care landscape. Barriers manifested differently depending on the existing organizational and technological capabilities. Data migration and acquisition costs were mentioned by many interviewees from both the supply and demand sides. Cloud technology posed fewer barriers to organizations with few installed on-premise systems that sought to either implement new pure cloud-based solutions or rely on a software as a service business model. On the contrary, organizations with relatively mature digital infrastructures and...
legacy systems faced hurdles to transition to cloud only solutions as they had to integrate existing systems and replace core infrastructures. Existing legacy systems were based on proprietary data structures and workflows, meaning that they could not simply be imported into the cloud. Instead, these organizations were more amenable to hybrid cloud solutions that relied on infrastructure as a service implementations:

The problem is, the cost of transition, if you’re talking about your patient administration system or your electronic health record, which is often the core bit of software in your health care organization, if they want to switch that out, it is a huge job, which is massively expensive and massively risky to do. [Participant 4, male, software vendor, United Kingdom]

In addition, implementers in particular raised the barriers associated with the need to change their cost structure with cloud technologies from capital up-front investments to a revenue model with recurring costs. This was perceived to be particularly problematic during the transitional period, when organizations were often running and paying for parallel systems:

In the short-term, you are inevitably paying more for the move towards cloud because you haven’t necessarily got rid of all of that other infrastructure as you make that transition. So, you’re now starting to pay for a revenue cost for your new cloud platform, but you’ve still got all of the cost of that other physical environment until you’re able to decommission. [Participant 16, female, research, United Kingdom]

Barriers not only related to cost but also to the organizational capabilities to adopt cloud solutions. Here, a lack of existing knowledge and skills in organizations to deploy and exploit cloud functionality was an important rate-limiting step. For instance, organizations frequently lacked implementation and migration skills:

In order to move things securely to the cloud either to implement brand new or to do a migration, you know, you need to have a certain degree of skill, knowledge, capability in order to do that... [Participant 4, male, software vendor, United Kingdom]

Existing technical skills and capabilities also played an important role in maximizing the benefits of cloud functionality once it was implemented. Here, participants stated that many health care organizations lacked the knowledge and skills needed to work with advanced large-scale data analytics and therefore struggled to optimize the use of cloud infrastructures through artificial intelligence and ML:

There’s a step that still needs to happen in the healthcare space, which is around just understanding what the analytics is. [Participant 19, male, software vendor, United Kingdom]

Other barriers inhibiting uptake of cloud technologies in health care organizations included the changes in organizational culture required to transition to externally hosted systems and new modalities of accessing critical services. This was seen as particularly problematic for a risk-averse sector such as health care. For example, some participants mentioned that organizations that were skeptical about implementing cloud technology feared a loss of control if they migrated their IT systems to external service providers. In addition, there was apprehension about the reliability of the cloud and telecommunication infrastructure to deliver critical services, which manifested in the perceived need to fall back on on-premise IT services as contingency measures for critical services:

Traditionally, IT departments in [provider organizations], you have your server, you have your software on it, and they manage that. It makes them slightly uneasy if it’s out there in a cloud and it’s not something that they have control of. [Participant 4, male, software vendor, United Kingdom]

Others stated that moving to cloud technology threatened established organizational hierarchies, particularly when sharing data across organizations. Health care settings were often not used to working across organizational boundaries. Cloud services challenged the traditional conception of organizations as autonomous entities and posed dilemmas in relation to information governance:

It’s about [organizations] having to give up something to be part of a bigger collaboration. [Participant 21, male, health care provider, United Kingdom]

Infrastructural Changes Not Immediately Visible to Frontline Users, Resulting in Lack of Clinical Pull

Although the organizational benefits of a wide range of cloud-based functions were visible and the case for organizational process and workload improvements could be made relatively easily by suppliers and system implementers, there was a perceived gap in visible benefits for frontline clinicians and patients. This presented a key barrier to the wider uptake of some cloud-based services as end users need to be on board for organizational changes to be implemented effectively:

An organizational imperative has to pass the challenge of the clinicians’ view of what is important and vice versa. The clinicians’ view of what is important has to pass the challenge of the gatekeepers in terms of organization, of funding, of development, service development, building development. [Participant 18, male, software vendor, United Kingdom]

The underlying issue was the invisibility of digital infrastructures for those at the frontline, who mainly experienced benefits through the exploitation and optimization of these infrastructures once they were in place:

The people who are going to be using the technology, the people who are going to be using the insights from the analytics, the people who will be experiencing the change in process, they are almost don’t really, it might sound harsh but...in the heat of the moment they almost don’t really care about is it cloud enabled? What is the infrastructure? What’s going
This lack of immediately visible benefits for end users combined with concerns surrounding privacy and security and the handling of sensitive data led to a lack of active user pull for cloud technologies in health care. It also presented challenges for suppliers as they had to satisfy a range of demands surrounding not only business processes but also clinical utility:

There’s some particular challenges, how do you deal with the privacy aspects of the data and satisfy the concerns that data contributors and data custodians have, and then how do you accommodate for this enormous diversity within the user community in terms of how they use data and importantly how they get beyond very simple table analytics views of data into something that is more problematic, and how do you find a way for those outputs, those research outputs, to make their way back into clinical utility. [Participant 7, male, software vendor, United Kingdom]

Despite these uneven perceptions, we also observed that during the COVID-19 crisis, clinical benefits and experiences of cloud technology became more common and thus immediately visible as remote consultations, remote working, data storage, and automation (eg, through chatbots) increasingly became a necessity:

Overnight we did see this huge uptick in the amount of telehealth, and that was only possible because of cloud there to support it. [Participant 22, female, cloud vendor, United States]

Visions of the Future Cloud Vendor Ecosystem

Innovators, implementers, academics, and cloud vendors agreed on a vision characterized by a hybrid cloud-enabled ecosystem of applications where software suppliers rely on a combination of on-premise systems and cloud integration with a large cloud provider. For software suppliers, integrating with a cloud platform meant that they could quickly and cost-effectively scale up and scale down their products as required. This, in turn, was perceived to translate to lower risk and more efficient costing for health providers surrounding the trialing of new services:

For us, the main use cases are around working with a platform that allows us to quickly and cheaply get our product out into market...we don’t need to invest huge amounts of time and people in developing things that are already out there...We can manage and maintain one environment, rather than having to think about how do you easily deploy and support, maintain, you know, 10, 20, 100 different customers, and the intricacies of deploying our app at every single customer site. We only have to think about one location. [Participant 20, male, software vendor, United Kingdom]

However, the participants (in particular, implementers and software developers) also flagged the challenges and risks in terms of interoperability between different platform providers and integration between software vendors and cloud vendors. Innovators and system implementers voiced their expectations for interoperability standards and for cloud providers to open up application programming interfaces. However, opening up application programming interfaces and standardizing key functions was not always in line with legacy providers’ commercial interests, which were typically based on retaining users within their platforms. Therefore, innovators in the software industry and implementers within health care organizations called for national regulations specifying interoperability standards to avoid vendor lock-in as this would allow for integration between systems and improve data portability. A lack of interoperability standards was viewed as inhibiting the development of a vibrant cloud ecosystem:

These regulatory bodies inside each of the governments would say the same thing, because that is the way to drive adoption of new technologies, forcing the new adoption, not rewriting everything, that’s out of the question, but forcing for the benefit of all. I think this is how you’re going to be having a government that is strong on that. [Participant 9, male, cloud vendor, Europe]

Discussion

Principal Findings

Although the participants perceived clear drivers for the use of cloud technology in health care settings, particularly in relation to collaboration and workload efficiencies, barriers to progress included data migration costs and skill gaps within health care organizations to support implementation. This was exacerbated by the perceived cultural shift required to move to externally hosted services, challenging entrenched organizational ways of working and the need to reorganize existing cost structures. Frontline users, particularly those lacking technical expertise, were not directly concerned with the benefits associated with cloud-based infrastructures, which resulted in a lack of user pull in organizations seeking to change their technological infrastructures. However, the pressures of the COVID-19 pandemic and the stronger need for remote working arrangements made various critical cloud services visible. Central regulations and mandated interoperability standards were viewed as a key priority to foster innovation and reduce the risk of vendor lock-in.

Integration of Findings With the Current Literature

Our study confirms findings in other sectors that highlight that, despite the potential benefits, the move to cloud-based technologies in organizations necessitates cultural shifts from established ways of working and administering systems [26]. Therefore, it needs to be viewed as a complex sociotechnical transformation process, requiring not only technological but also socio-organizational changes to maximize the potential of cloud technologies [27]. Here, changes in organizational business models and technological infrastructures associated with cloud technology are likely to affect existing ways of working and organizational functioning as a whole [28]. Therefore, a key area of focus needs to be the effective
integration and embedding of new infrastructures with the installed base of existing technologies and socio-organizational structures and practices [4]. Barriers associated with data migration to cloud-based solutions are well documented in the literature [29], but our work also points to differences between digitally mature organizations with established installed technological systems (requiring more fundamental changes to the installed technological base) and those organizations that do not have established technological infrastructures, where data migration is likely to be less of an issue.

There is an asymmetry in the way system implementers, clinicians, and patients perceive and understand the benefits of the cloud, particularly when it comes to advanced functions such as ML and data-driven functionalities, which results in a lack of strong user pull [30]. User pull to implement cloud technologies within organizations is critical, especially in public service sectors [31]. Here, user attitudes and expectations toward technology can have a direct impact on adoption patterns [32]. A lack of perceived direct benefits as well as skepticism and concerns (most notably, perceived security, trust, and privacy issues) can result in negative attitudes toward a technology and lead to abandonment [33]. There are now growing calls for transparency and accountability of how personal data are used within cloud-based systems without compromising privacy and security [34]. Medical research is a key area where clinical data are considered immensely valuable but where handling of sensitive data is of utmost importance. This issue intersects not only with privacy and security but also with growing interest across industry and academia on trustworthy, fair, and ethical use of big data and algorithmic technologies [20,35,36]. Therefore, it is critical for organizations promoting the use of cloud technology to place emphasis on active engagement with users and rigorously engage with debates about privacy, ethics, and security taking place in academic and public forums [37].

The move to cloud technologies in health care presents a disruptive innovation for the market [30,38], which inevitably results in tensions and trade-offs between conflicting agendas and interests. In this study, we observed that points of friction related to the integration of different building blocks and interoperability between competing platforms. These challenges resonate with previous studies in information systems, which highlight ongoing tensions between requirements for standardization and the flexible and cost-effective operation of systems [39-41].

**Implications for Policy, Practice, and Research**

Among the key challenges voiced by our informants were the lack of installed capacity and technical skills, the cost of migration, and the need for investment restructuring. As a result, organizations that still rely on on-premise IT infrastructure and software see hybrid cloud solutions as a way forward. There is now a need to support the development of such hybrid structures and map potential integration and migration pathways to help implementing organizations envisage new information infrastructure constellations. This needs to be supported by active efforts to address the existing skill gap in cloud computing and digital transformation expertise in the health sector [42]. This will also help ensure that advanced cloud functions such as ML are effectively exploited.

Strategic decision makers need to recognize the need to view the implementation of cloud-based systems as a major digital transformation of services to promote cloud first policy in health care settings. Therefore, implementations need to be supported not only by technological capability but also by change management expertise and continuous stakeholder engagement.

Our work highlights divergent views and expectations among various stakeholder groups in relation to interoperability. These are highly contingent upon political-economic contexts as interoperability standards are not always centrally mandated across countries. Innovators and system implementers in particular raised the need to regulate the emerging cloud ecosystem through the development of interoperability standards. Adding to the risk of developing solutions for a particular vendor is poor integration between competing platforms. A clear policy recommendation to address this challenge is the central mandate for interoperability standards, with the United States being a case for reference, but these need to be flexible to respond to emerging needs and other disruptive innovations that are likely to emerge. Of central importance will be the need for trustworthy entities and tools for responsible use of sensitive data, developing mechanisms for ensuring ethical and transparent use for medical research without compromising patients’ privacy and integrity.

**Strengths and Limitations**

We gained insights into the opportunities and challenges in the emerging area of cloud technology implementation in health care settings by consulting a range of perspectives. We deliberately sampled implementers, customers, academics, and vendors to explore experiences and insights from a range of settings. However, this may have been at the expense of breadth. For example, consumer and customer perspectives were underrepresented in our sample, and we did not consult the range of immediate frontline users of technologies or legal and privacy experts. Our sample also consisted mainly of cloud enthusiasts. Nevertheless, our study points to various user-facing issues such as adoption, use, concerns, and invisibility of functions, which we assessed indirectly through respondents working in close contact with users. Further empirical work with clinicians, lawyers, and privacy experts arises as a pertinent avenue of research.

Our themes provide a helpful guide for conducting future in-depth work as we have illustrated an overview of tensions. In addition, we would also have liked a broader representation of international settings (as 18/23, 78% of participants in our sample were based in the United Kingdom). Our current sample consisted mainly of participants from North America and Europe (France, Finland, and the United Kingdom). Future work should build on our findings seeking to explore how different geographies, including low- and middle-income countries, have approached the area and how challenges vary across different core infrastructures, levels of digital maturity, and health system organization.
Conclusions
Although cloud technologies promise to deliver a range of technical capabilities, they are unevenly applied across health care settings depending on organizational contexts and existing infrastructures. In the wake of the pandemic, cloud technologies have become vital to support everyday collaboration for clinicians, remote health delivery, and other operational functions, which has considerably driven the adoption of the cloud. Going forward, cloud implementation needs to be viewed as disruptive organizational change initiatives facilitated by national initiatives to promote interoperability for a vibrant cloud ecosystem. Areas that may lend themselves to such work may include patient-facing technologies, where cloud providers are already established, and health and social care integration, where limited existing health information infrastructures may reduce barriers associated with integration or migration. This will also need to involve engaging in public discourse about cost, risk, and trust (or lack thereof) in cloud platforms regarding the handling of sensitive data, privacy, security, and ethics.

Acknowledgments
The authors gratefully acknowledge all the participants’ time and input. They also thank Scott Watson for his thoughtful comments on the paper and Rachel Dunscombe and John Halamka for helping with the recruitment of participants. This work was funded by a Scottish Government Chief Scientist Research Grant. The views expressed are those of the authors.

Conflicts of Interest
None declared.

References
5. The history of cloud computing. Server Mania. URL: https://blog.servermania.com/the-history-of-cloud-computing/ [accessed 2021-12-14]
14. Cloud First is here to stay. GOV.UK. URL: https://technology.blog.gov.uk/2019/10/31/cloud-first-is-here-to-stay/ [accessed 2021-12-14]


34. Securing the E-Health cloud. ACM. URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=7C1DC751764B9D7CDD09D023304FCA3?doi=10.1.1.717.462&rep=rep1&type=pdf [accessed 2021-12-14]


38. Cloud computing as a disruptive technology. Cloud Academy. URL: https://cloudacademy.com/blog/disruptive-technology-and-cloud-computing/ [accessed 2021-12-14]


41. Understanding infrastructure: dynamics, tensions, and design. University of Michigan Library. URL: https://deepblue.lib.umich.edu/handle/2027.42/49353 [accessed 2021-12-14]

Abbreviations

**IT:** information technology  
**ML:** machine learning  
**TPOM:** Technology, People, Organizations, and Macroenvironmental factors
Comparing International Experiences With Electronic Health Records Among Emergency Medicine Physicians in the United States and Norway: Semistructured Interview Study

Gracie Garcia¹, BA; Christopher Crenner¹, MD, PhD
Department of History and Philosophy of Medicine, University of Kansas School of Medicine, Kansas City, KS, United States

Corresponding Author:
Gracie Garcia, BA
Department of History and Philosophy of Medicine
University of Kansas School of Medicine
3901 Rainbow Blvd
Kansas City, KS, 66160
United States
Phone: 1 6209514171
Email: grace.c.garcia1@gmail.com

Abstract

Background: The variability in physicians’ attitudes regarding electronic health records (EHRs) is widely recognized. Both human and technological factors contribute to user satisfaction. This exploratory study considers these variables by comparing emergency medicine physician experiences with EHRs in the United States and Norway.

Objective: This study is unique as it aims to compare individual experiences with EHRs. It creates an opportunity to expand perspective, challenge the unknown, and explore how this technology affects clinicians globally. Research often highlights the challenge that health information technology has created for users: Are the negative consequences of this technology shared among countries? Does it affect medical practice? What determines user satisfaction? Can this be measured internationally? Do specific factors account for similarities or differences? This study begins by investigating these questions by comparing cohort experiences. Fundamental differences between nations will also be addressed.

Methods: We used semistructured, participant-driven, in-depth interviews (N=12) for data collection in conjunction with ethnographic observations. The conversations were recorded and transcribed. Texts were then analyzed using NVivo software (QSR International) to develop codes for direct comparison among countries. Comprehensive understanding of the data required triangulation, specifically using thematic and interpretive phenomenological analysis. Narrative analysis ensured appropriate context of the NVivo (QSR International) query results.

Results: Each interview resulted in mixed discussions regarding the benefits and disadvantages of EHRs. All the physicians recognized health care’s dependence on this technology. In Norway, physicians perceived more benefits compared with those based in the United States. Americans reported fewer benefits and disproportionately high disadvantages. Both cohorts believed that EHRs have increased user workload. However, this was mentioned 2.6 times more frequently by Americans (United States [n=40] vs Norway [n=15]). Financial influences regarding health information technology use were of great concern for American physicians but rarely mentioned among Norwegian physicians (United States [n=37] vs Norway [n=6]). Technology dysfunctions were the most common complaint from Norwegian physicians. Participants from each country noted increased frustration among older colleagues.

Conclusions: Despite differences spanning geographical, organizational, and cultural boundaries, much is to be learned by comparing individual experiences. Both cohorts experienced EHR-related frustrations, although etiology differed. The overall number of complaints was significantly higher among American physicians. This study augments the idea that policy, regulation, and administration have compelling influence on user experience. Global EHR optimization requires additional investigation, and these results help to establish a foundation for future research.

(JMIR Hum Factors 2022;9(1):e28762) doi:10.2196/28762
KEYWORDS
electronic health records; electronic medical records; health information technology; health information exchange; health policy; international; emergency medicine; medical informatics; meaningful use; burnout

Introduction

Background

Correlations between electronic health records (EHRs) and physician frustrations have been well described throughout informatics literature. The phenomenon of high user dissatisfaction is often attributed to increased administrative requirements, decreased face-to-face patient time, information overload, and limited interoperability [1-12]. This technology has been analyzed on both local and global scales [13-17]; however, few studies have compared users from different countries who practice in parallel clinical settings. Our study compares emergency medicine (EM) physician experiences with EHRs in the United States and Norway. In addition, societal and cultural differences are carefully considered while analyzing components that may affect user satisfaction.

Health information technology (HIT) is used in many countries, but deployment of EHRs vary [18-23]. Despite global HIT use, the United States is perhaps the most prominent generator of informatics research that emphasizes the shortcomings of this technology. Currently, there are few studies that consider or compare international EHR experiences. However, a recent study by Downing et al [24] found that even when using the same vendor (Epic Systems), American physicians had significantly longer documentation and were less likely to report satisfaction or improved work efficiency compared with those in Australia and Singapore. Our results contribute to this small body of research.

Objectives and Measured Outcomes

This study considers factors that contribute to EHR user satisfaction by comparing individuals with similar professional responsibilities in different national contexts. Are the negative consequences of this technology shared among countries? Does it affect medical practice? What determines user satisfaction? Can this be measured between international cohorts? Are there specific factors that account for similarities or differences? This study explores these questions while considering influential variables from a sociopolitical–technological context.

Primary outcomes include the overall EHR experience and specific opinions within each cohort. This was achieved by conducting structured interviews in the hospital and observing behaviors within the physician’s typical working environment. Thematic analysis allowed the quantification and comparison of common topics. Notable differences may help to identify targeted solutions for HIT optimization. For example, if users from each cohort believe that software interfaces are challenging to use, it could indicate that technology-specific factors (understanding and using computers) significantly increase frustration. On the other hand, differences may identify solutions that may have otherwise been overlooked.

Secondary outcomes assess participant responses within a sociocultural context, as HIT infrastructure differs among countries [16-23,25]. Previous research shows that successful EHR use is greatly influenced by social and governmental constructs [9,24,26,27]. A general understanding of the current HIT status and health care infrastructure in the United States and Norway supports the interpretation of the data. We briefly discuss this information before proceeding.

United States: Emergency Care and Current EHR Status

Since the Emergency Medical Treatment and Labor Act of 1986, hospitals must provide consultation, screening examination, ancillary testing, and stabilization of anyone concerned with a life-threatening condition, regardless of their ability to pay [28-30]. Patients are evaluated in the emergency department (ED) after arriving via ambulance, private vehicle, or walking in. Physicians who staff the ED receive formal EM training by completing an EM residency for 3-4 years following medical school [31]. Although the ED functions as a hospital’s gatekeeper, studies show that only a small number of ED visits result in admission [32]. National increase in low-acuity ED patient volume has been attributed to multiple factors including rising health care costs, primary care shortages, and lack of access to after-hour care [33-37].

Integration of technology and health care started in 2004 with the establishment of the Office of the National Coordinator for HIT, but widespread EHR use did not occur until after the HIT for Economic and Clinical Health Act was passed in 2009 [38]. This legislation provided monetary incentives for government-certified EHR adoption and implementation [39]. Pressure for rapid health care digitization generated numerous unintended consequences including industrial arms race that many policy makers did not consider [40-43]. As of 2017, 96% of hospitals in the United States had implemented technology certified by the Department of Health and Human Services [44].

The 21st Century Cures Act prohibits companies and organizations from intentionally restricting health information exchange (HIE) capabilities for monetary benefit [45]. Nevertheless, evidence indicates that information blocking still occurs in the United States [46,47], and a 2018 Report to Congress showed only 51% of hospital physicians had electronic access to necessary patient information from other facilities at the point of care [44]. The private sector has been meeting interoperability demands as evidenced by programs like Epic’s Care Everywhere [27]. However, the extent of clinical data availability is dependent on the participating facilities [48]. In 2018, the US Department of Veterans Affairs (VA) announced partnership with Cerner, an EHR company that will eventually be the sole vendor to all VA facilities that serve military populations [49].

Norway: Emergency Care and Current EHR Status

Inpatient and specialist care are provided by state-owned hospitals and managed by 4 geographically distinct government subdivisions known as Regional Health Authorities [50,51]. A
total of 428 local municipalities are responsible for supplying primary care including after-hour access [50]. Municipalities have urgent care centers with on-call physicians (legevakt) [50]. The ED or acute receiving area (akuttrett) is only accessible via ambulance or physician referrals [50,52]. The department is traditionally staffed by internal medicine, neurology, orthopedics, and surgery physicians [53]; however, EM was recently recognized as an independent specialty in Norway in 2017 [52]. Historically, ambulance and other health personnel would communicate with hospitals to determine the most appropriate inpatient specialty service to receive the patient upon arrival [50].

Medical records from hospitals and outpatient facilities are not integrated, but messaging systems embedded within EHR software allow providers to collaborate [51]. In 2008, the government recognized the interoperability needs and launched a national HIE platform in 2012 known as Core Journal (Kjernejournalen) [54]. This gives all Norwegian physicians access to critical patient information, regardless of where previous treatment was provided [51,54]. It includes data necessary to prevent unfavorable outcomes that may be difficult to obtain during emergency situations such as severe allergies, ongoing treatments (eg, dialysis), rare serious conditions (eg, hemophilia), and medications dispensed at any Norwegian pharmacy [55]. Research shows that the most used function is the pharmaceutical tracking tool as it provides up-to-date medication information without additional manual data-entry requirements from physicians [56,57].

In 2013, the Directorate of Health recommended the integration of all eHealth and developed the initiative One Patient–One Record (Én innbygger–Én journal) [58,59]. In 2019, a US $296 million contract was signed with an American EHR company (Epic Systems) to eventually function as the nation’s sole HIT supplier [13,60]. The pilot program Health Platform (Helseplatformen) is scheduled to launch during the spring of 2022 in Central Norway, 1 of the 4 Regional Health Authorities [61]. Current studies indicate optimistic expectations mixed with concern as protected health information will eventually be exchanged across administrative, geographical, and institutional boundaries [62]. Regional governments created consensus groups comprised of health care professionals from >80 municipalities that are involved in software configuration and design [63]. After implementation, community physicians and analysts will continue to optimize the functionality for regional and practice needs, whereas Epic Systems will be involved to a lesser extent [13].

**Methods**

**Participants and Setting**

This study was conducted at the University of Kansas Medical Center (KUMC) in Kansas City, Kansas, and at the Akershus University Hospital in the Lørenskog municipality outside of Oslo. Bed capacity at each hospital was approximately 1000 beds [58,59]. Recruitment emails were sent to physicians involved in acute care at these facilities. In the United States, participants were board-certified EM physicians, whereas in Norway, participants were surgeons who provide services within the akuttrett. A total of 12 interviews were conducted, 6 (50%) at each location. Average conversation lengths were 39.1 (SD 15.8) minutes.

**Data Collection**

Data collection included face-to-face semistructured interviews and environmental observations. This was possible by conducting each interview on site at the hospitals. Participants were willing to show the typical documentation and clinical workflow to the interviewer (GG). This was essential when collecting Norwegian data, as the interviewer had no previous first-hand experience with this health care system. This provided context when participants referred to specifics of the EHR. Without this background the contextual understanding of participants’ answers would have been severely limited. All the interviews were conducted in English, as all the participants were proficient in this language. Conversations were audio-recorded on a passcode-protected device and then transcribed for further analysis. Privacy was retained by deidentifying the participants. After obtaining written informed consent, standardized questions were used to obtain the following information from each participant: (1) demographics, (2) cultural and individual values, (3) individual comfort with general technology, (4) previous record experiences (electronic or paper), (5) observations of colleagues regarding EHR use, (6) individual attitudes toward EHR at current facility, (7) perceived usability (intuitive interfaces, software functionality, interoperability, workflow efficiencies, and centralized data repository), and (8) how the technology has shaped individual practice.

Follow-up questions varied based on individual responses. Participants were also asked about their knowledge, opinions, or questions regarding the other cohort’s electronic health care infrastructure. Natural conversation flow permitted additional discussion, allowing deeper exploration of ideas as they appeared organically. Additional questions developed throughout data collection were based on previous participant answers and cumulative observations. For example, US interviews were completed first and the responses involved specific negative consequences of the EHR without prompting. If these topics were never mentioned by the Norwegians, the interviewer inquired about them directly at the end of the discussion.

To conclude each interview, participants were asked if they had specific questions for the physicians in the other country. Following data collection, questions and answers were distributed to participants in addition to the background information on each country’s health care system. This allowed deeper understanding of individual perceptions while generating rich discussion. In addition, participants in Norway were explicitly asked about Kjernejournalen use. This study was reviewed by and received institutional review board approval from KUMC while abiding by the General Data Protection Regulation.

**Analysis**

The US interviews were completed first, followed by interviews in Norway. Using grounded theory, themes emerged and evolved
throughout the entire data collection process. As no single method captured the complexities of these data, analysis triangulation was necessary. First, transcripts underwent numerous thematic analyses to identify patterns between the cohorts. This was the initial formal approach to derive meaning from the vast and rich collected data. Similar to grounded theory, this exploratory methodology allows continuous hypothesis development throughout analysis progression. Narrative analysis was conducted to provide further insight into the mindset, perspectives, and attitudes toward EHRs. In addition, direct quotes were used to support the findings and may help the reader appreciate the nuances of the social context and emotion.

Early in the analysis process, 2 broad themes were identified—perceived EHR benefits and perceived EHR disadvantages. To gain deeper understanding of the data, interpretive phenomenological analysis and simple content analysis were used. Both methods aid in succinctly summarizing concepts based on individual experiences while providing some quantitative comparison. These techniques paired with the NVivo software (QSR International) helped to distinguish conceptual patterns between the cohorts, and ultimately resulted in the construction of the following 4 main code groups: US perceived EHR benefits, Norway perceived EHR benefits, US perceived EHR disadvantages, and Norway perceived EHR disadvantages.

Transcriptions were analyzed using the NVivo software (QSR International). As the perceived EHR benefits or disadvantages were found within the text, they were assigned to 1 of the 4 code groups based on context and cohort. The NVivo word frequency and query search functions were used to generate categories within the encoded text to enrich the results. The software allowed searches to include exact word matches, stemmed words, and synonyms. The search criteria details are presented in Textboxes 1 and 2. Identical queries regarding perceived EHR benefits and disadvantages were conducted for both cohorts. Query results were analyzed and refined to ensure that the terms were not taken out of context. The total number of results for each group was tabulated and compared. This is displayed in Figure 1. Comparing the categorical patterns provides concrete examples of varying priorities, opinions, and perspectives from the 2 cohorts. In addition, it examines the advantages and flaws of HIT implementation within each health care system.


**Category and search criteria (include exact matches, stemmed words, and synonyms)**

- Patient safety and improved care: safety, benefit, care, improve, alert, allergy, interaction, medication, automated, error, writing, legible, and mistakes
- Access to useful clinical information: accessibility, information, view, records, journal, chart, review, report, previous, tracking, results, history, important, critical, clinical, diagnosis, remote, exchange, facility, interoperable, capability, cloud, electronic, time, and speed
- Data organization: organization, sort, filter, search, usability, function, central, record, history, chart, journal, ease, efficient, and review
- Enhanced communication: communication, interaction, order, results, review, chart, record, journal, information, patient, encounter, and clarification

Textbox 2. Categories of perceived electronic health record disadvantages.

**Category and search criteria (include exact matches, stemmed words, and synonyms)**

- Excessive or irrelevant data: excess, irrelevant, overload, quantity, redundant, limit, volume, amount, organize, filter, lost, search, data, clinical, benefit, and documentation
- Poor interoperability: interoperability, access, view, restrict, facility, exchange, data, information, chart, journal, record, outside, cloud, capability, hospital, and clinic
- Increased workload: work, workload, time, hour, administrative, requirement, documentation, efficient, amount, burden, click, task, and clerical
- Software complexities: software, complex, interface, intuitive, user, difficult, friendly, usability, navigate, understand, function, options, programs, system, load, slow, lag, ease, options, orders, run, and technology
- Hardware malfunctions: hardware, malfunction, crash, process, failure, update, IT, program, develop, technology, support, computer, device, speed, and paper
- Financial influence: financial, money, reimbursement, billing, profit, cost, incentive, code, dollars, relative value unit or RVU, regulation, mandate, clinical, value, price, payment, and business
This process was conducted by the interviewer for retained consistency while considering abstract factors including nonverbal communication, clinical environment, and cultural norms. This technique was repeatedly used to explore topic relationships, consider causality, and help find thematic saturation within the populations.

**Results**

**Overview**

All the participants described both pros and cons of their experience with the EHRs and both groups agreed that modern medicine is heavily dependent on this technology. In general, Norwegian physicians had a slight propensity to report benefits (62 total perceived benefits reported) compared with disadvantages (59 total perceived disadvantages reported). In contrast, the American cohort frequently expressed unfavorable perceptions, reporting 145 total perceived disadvantages and only 47 total perceived benefits. These results are summarized in Figure 1.

**Perceived Benefits**

Access to relevant patient information was the most commonly reported benefit in both countries. This included viewing previous diagnostic studies, clinical notes, and laboratory results. The Norwegian physicians were 1.7 times more likely to refer to these specific benefits (Norway [n=33] vs United States [n=20]). When American physicians mentioned this tool, they often also noted significant limitations owing to poor interoperability between competing HIT supply companies and health care facilities. A commonly perceived positive EHR outcome in both cohorts was improved patient safety. The results were moderately comparable between the 2 countries with American physicians referencing patient safety 20 times and Norwegian physicians referencing patient safety 15 times. An example that was frequently mentioned by participants was the automated alerts about patient allergies or drug–drug interactions. Many also believed that it has decreased unnecessary errors caused by illegible handwriting.

**Perceived Disadvantages**

In general, there was a much broader range of topics related to perceived disadvantages when compared with benefits. The belief that EHRs have increased physicians’ workload was common to both cohorts. However, this was mentioned 2.6 times more frequently by the Americans (United States [n=40] vs Norway [n=15]). The most reported disadvantage was how increased clerical work detracted from efficiency. American physicians also discussed that they believe the required documentation has minimal, if any, clinical utility.
In Norway, the 2 most frequently discussed disadvantages of EHRs included software complexities (Norway [n=36] vs United States [n=25]) and hardware malfunctions (Norway [n=15] vs United States [n=7]). Every other disadvantage category was more common among the US cohort. In addition to increased workload, other categories included excessive and irrelevant data (United States [n=25] vs Norway [n=12]) and poor interoperability (United States [n=34] vs Norway [n=14]). The most significant difference between the cohorts was regarding the financial influence of the EHRs (United States [n=37] vs Norway [n=6]). Each American physician expressed without prompting that the primary purpose of EHRs within the United States is for billing rather than to improve patient care. This was often attributed to competing business models among HIT suppliers, insurance companies, and hospital administrations.

One disadvantage exclusive to the American cohort involved the legal implications of the EHRs. The interviewer never initiated this topic, yet it was brought up by half of the American participants. They strongly believed that the normalization of defensive medicine is a result of the society’s legal climate. Despite possessing adequate medical training and clinical judgment, clinicians often feel compelled to order extensive workups to protect themselves from future prosecution. In addition, these physicians mentioned that redundant testing is routinely performed because of limited HIE among surrounding health care facilities.

**Additional Observations**

An interesting observation shared by both the cohorts was that their older colleagues expressed higher levels of EHR-related frustration. This was mentioned by 9 among all 12 participants—6 Norwegians and 3 Americans. These 9 individuals self-reported that they felt proficient in using technology but did not believe it influenced their own opinion of EHRs.

In addition, both Norwegians and Americans believed that the rapid processing speeds of personal devices may contribute to unrealistic EHR performance expectations. Many realized that top information technology developers are recruited to sectors outside of health care; however, they believed that usability would improve if companies such as Apple or Google developed the software:

> It’s very hard to keep up with ever-changing new technology. As you get older you don’t have the stamina. Programs may also seem frustrating because they don’t run as quickly as most of our personal devices. [Norway, participant 4]

At the start of each interview, the physicians were asked how cultural values influenced personal beliefs or medical practice, as other studies have described health care systems as a reflection of national ideals [26]. This question was intended to highlight nuanced variables that exist when comparing dissimilar populations. Unsurprisingly, the participant responses revealed differing values between the countries. Responses were not superior or inferior, just different. When describing how cultural values influence their current practice of medicine, the American participants used words such as *help, kind, and caring,*

Common Norwegian terms included *open-minded, equality,* and *empathy.* Although these results have limited application in determining EHR satisfaction, it reinforces the importance of cultural context when developing solutions for specific populations.

**Discussion**

Adding quantitative values to our qualitative analysis creates an overt visualization of the differences between EHR users in both countries. We have provided a more comprehensive exploration of the influencing factors.

**Clerical Burden and Reimbursement**

Increased administrative tasks that yield minimal patient benefit created frustration for all physicians; however, it was significantly higher among US participants. In Norway, physicians must include appropriate diagnosis or procedure codes for hospital reimbursement using the Diagnosis Related Groups system, which includes approximately 980 codes [64]. These codes generate approximately 50% of the hospital revenue, with the remaining financed from fixed government payments [65]. Norwegians are skeptical of potential changes following the national implementation of Epic Systems. The participants voiced concern regarding slowly evolving into an American health care model. Some Norwegian participants first noticed this shift after hospital reimbursement became partially integrated with diagnosis codes:

> With new public management reform within the last 30 years, we have also noticed health care has changed to suit their needs. Most Norwegian physicians are attentive and oppose this. We also have the union (Norwegian Medical Association) who oppose it. It isn’t in our immediate power to change those things and they must come from a higher level. [Norway, participant 6]

American reimbursement is complex owing to a multi-payer system that includes government agencies, insurance companies, health maintenance organizations, employers, and individual patients [66]. Although many countries use the International Classification of Disease, the United States is one of the few countries that use it for both diagnosis and billing, while including more than 90,000 codes [67]. Compared with Norway, the United States uses the entire medical record for reimbursement. The billing level is determined by the quantity of the documented elements within each note section (ie, history of present illness, review of systems, and physical exam) with more elements correlating with higher billing levels, resulting in increased reimbursement [68]. Physicians must also filter through long, redundant, and confusing lists of diagnoses to choose the most detailed option [24]. Another form of reimbursement, relative value units, is also extracted from the EHR. These are based on >8000 procedure codes extrapolated to measure physician productivity, which are then used to determine department or individual reimbursement [66,69]: 1 physician shared that for the last 2 years, 25% of their salaries depended on the individual relative value units generated.
Institutions now look at emergency departments as revenue generators. We cost society more and in the end the patient loses directly and indirectly. [United States, participant 3]

American documentation tends to be 4 times longer than that of other countries, without offering any additional clinical information [48]. The position of medical scribes (nonclinical personnel who are trained to provide documentation assistance and workflow support) was created as a possible solution to this problem. Research demonstrates that scribes are valued team members and improve provider satisfaction [70]. This sentiment was echoed by American physicians, whereas Norwegians were unfamiliar with this occupation. Gardner et al [1] showed great variety of scribe use among American specialties, with the highest use among EM physicians. This study also found that working with scribes reduced the odds of burnout by approximately 40%. They hypothesized that it was not higher because scribes are not qualified to complete certain time-consuming but physician-specific electronic tasks (eg, medication orders and in-basket management) [1]:

We are so opposed to these tasks that steal time we could otherwise use for clinical work. Don’t you think that having a scribe is just a waste of resources? Do individual physicians actually generate enough data on a single patient that they need a scribe to help complete the documentation? [Norway, participant 6]

Burnout

Multiple studies within the United States indicate that HIT creates undue physician burden and there is considerable correlation between high EHR frustration and burnout [1,4,7,24,43,71-76]. The 2018 National Physician Poll produced powerful data regarding how this technology affects American physicians. Only 8% of participants believed that the primary purpose of documentation is clinical, whereas 71% agreed that it significantly contributes to burnout [77]. Our study supports this argument, as many American physicians cited EHRs as a significant cause of burnout. However, these individuals clarified that it is only a single contributor to a complex and multifactorial issue:

I don’t necessarily think the electronic aspect of EHRs are what makes them so frustrating, but rather the need of documenting in excess. When you have to do such complicated things against your will and without patient benefit, it adds to burnout rates. [United States, participant 6]

In Norway, burnout was never mentioned spontaneously and eventually the interviewer was required to ask about it explicitly. Norwegians attributed burnout to perceived job demands, societal expectations, and degree of colleague support, which is consistent with other Norwegian studies [78-80]. EHRs were never mentioned as a source of burnout. Since 1993, Norway has conducted extensive research aimed at improving physicians’ health, working conditions, and quality of life [81]. Despite burnout being less prevalent, Norway has established proactive prevention initiatives. An example is a self-referral physician counseling program and treatment facility (Villa Sana) designed to enhance coping skills and reduce emotional exhaustion [82,83].

West et al [76] considered factors that contribute to physician burnout from a global perspective. In doing so, they highlighted a previous Norwegian study that found no significant difference in burnout between physicians and other professions [84]. However, in the same study, there was a significantly increased prevalence of burnout in the United States even after adjusting for work hours and other factors [76]. Another recent US study identified systemic issues contributing to EM physician burnout. Factors include EHR limitations, long work hours, substantial educational debt, intense clinical practice, high risk of litigation, circadian rhythm disruption, chronic fatigue, blame, and isolation as a result of poor outcomes, all within the confines of an environment with zero tolerance for mistakes [85]. Our study offers informal evidence that EHRs increase burnout risk in the United States but appear noncontributory in Norway.

Core Journal (Kjernejournalen)

Of the 6 Norwegian physicians, 5 used the Kjernejournalen at least multiple times per week. Most information required initial manual entry, which has created additional tasks for providers. Some participants also attributed slow processing speeds as a reason for their limited use. However, the Kjernejournalen software provides a function that was highly favored by all the Norwegian physicians in this study—the pharmaceutical tracking function. The Kjernejournalen connects with all the pharmacies in the country and updates automatically as prescriptions are filled [55,56]. This tool was favored as it provides useful information without increasing data entry responsibilities. When asked if the Core Journal has affected their medical practice, the first participant provided the following response:

I would say that one way is you can now see what is prescribed and if it has been collected. It is a more secure way of finding out what patients are really taking. [Norway, participant 1]

Overall, there were mixed feelings about the software among Norwegian physicians. In contrast, all the American physicians expressed their desire for something similar upon learning about the Core Journal. They were also interested in the pharmaceutical tracking function, specifically for narcotic medications. Of the 50 states, all except one (Missouri) have state-wide tracking software; however, communication between programs is limited [86]. In addition, the American cohort at the KUMC faces the unique challenge of working within a facility that is geographically located on the Kansas–Missouri state-line border.

Interoperability

In-depth conversation regarding EHR interoperability capabilities revealed significantly different experiences between the EDs in the 2 countries. In Norway, specialty care is confined to hospitals and allows EM physicians to easily view specialist or inpatient notes. However, primary care facilities are part of the private sector and use different EHRs. Hospital and primary physicians alike are able to access the Core Journal, which provides information regarding critical diagnosis and current

https://humanfactors.jmir.org/2022/1/e28762
prescriptions [55]. However, Norwegian participants indicated that emergency care was never impeded because of the inability to access primary care clinic notes. Instead, their frustration occurred when requesting imaging from distant facilities. Both cohorts reported needing outside records and imaging occasionally. All the physicians found this task to be annoying and time-consuming. In Norway, all radiologic studies can be electronically exchanged among health care systems throughout the country and sometimes require several phone calls. It was reported that this can take up to 20 minutes but is typically completed more quickly. American physicians noted that they can occasionally view outside imaging. However, this is often not available and scans have to be repeated.

Patients in the United States often receive both primary or specialist care in an outpatient clinic setting. Providers have limited access to patient information at the point of care if health care facilities use different HIT suppliers [12,87]. Over the past decade, laws have been passed with the goal of improving interoperability, but definitive legal parameters are yet to be firmly established [12,45,46]. HIE configuration decisions are typically dependent on the competing vendors and participating health care systems, with both parties having significant effect on user accessibility [48]. Vendors have capitalized on developing exchange capabilities as a product selling point [46]. Subsequently, there have been calls for stronger legislative regulation to improve transparency across health care facilities [41,45,46].

Although individual EHR suppliers have improved interoperability, substantial limitations persist [19]. For example, American physicians in this study discussed the Care Everywhere platform within Epic Systems that grants access to most outpatient documentation and laboratory results from another large hospital within Kansas City. However, this tool still omits numerous facilities and hospitals. US participants reported that electronic exchanges between unaffiliated health care facilities are either impossible or extremely cumbersome and time-consuming. A participant described the process used to request outside medical records and said that it could take hours to days to receive a fax that potentially contains critical information. Knowing that the information will not be available within their own shift, this participant typically makes these requests to benefit colleagues who are taking over patient care. American physicians also believed that redundant diagnostic tests are a direct result of limited interoperability that increases both patient risk and national health care expenses:

> We repeat so many x-rays, labs, and scans just because we can’t see what was done a day ago. There are deficits in care due to poor EHR interoperability. Today, in this emergency department, there will be an issue because they [outside EHRs] don’t communicate. [United States, participant 6]

The VA is a government-run national health care system that internally developed its own EHR software known as Veterans Information Systems and Technology Architecture [41]. Each US participant who mentioned past VA experiences recalled positive experience with this EHR. Although the participants described the software’s interface as cumbersome and rudimental, all of them commented about how it allowed them to provide more comprehensive care because of the ability to access all the pertinent information from any VA facility. Despite the recent contract with Cerner, it will likely take longer than 10 years to finalize the implementation of this software as the sole HIT supplier to all VA facilities [41,49].

**Legal Considerations**

Another burden unique to Americans is the extensive documentation for legal protection. A recent study showed that approximately 51% of EM physicians in the United States will be sued during their career despite appropriate medical management [88]. This was foreign to Norwegians who rely on the Norwegian Medical Association (NMA) for legal counsel and protection [89]. NMA also functions as a professional society and labor union that annually negotiates with the government on behalf of physicians regarding fair working conditions, compensation, and leave-time [89]. Nearly all Norwegian physicians are NMA members, whereas only 11.4% of American health care providers are unionized [90]. Explanations for low involvement include convoluted multi-payer systems, restrictive federal and state laws, and social stigma [90-92].

Defensive medicine is a normalized practice within US medicine. American EM physicians face approximately a 7.5% annual risk of litigation [93]. Consequently, excessive documentation becomes an essential burden to protect oneself from potential legal ramifications. This liability heavily influences medical decision-making, resulting in excessive workups and hospital admissions. A study of 824 physicians in the United States found that 93% of them reported regular practice of defensive medicine [94]. Of those, more than half of the EM physicians reported using computed tomography, magnetic resonance imaging, or radiography that was not clinically necessary [94].

Responses from the American participants correlated with these findings and many believed that improved interoperability between EHR systems could mitigate these practices while simultaneously decreasing physician litigation anxiety. American participants also noted numerous disadvantages associated with defensive medicine on a societal, patient, and health care provider perspectives; however, abandoning this practice puts the physician at an undeniable risk:

> A lawyer can go through and subpoena every keystroke made from the moment you enter the record, what is done before completing the note, and if you changed anything. We are humans and will make mistakes. If you type something wrong, it can potentially be used against you to criticize your medical judgment. If I have a learner (i.e., scribe or resident) who wrote something wrong and I change or delete it, that may be held against me. [United States, participant 6]

This is in stark contrast to the practices in Norway, where physicians pay a small percentage of their salaries to a collective pool within the NMA. If a patient is entitled to compensation, it comes from these funds. All Norwegian participants expressed...
that this was a fair and equitable process without many disadvantages, and one physician stated the following:

I am only concerned for malpractice because I am always concerned with doing the right thing for my patient. I am not concerned about repercussions for making a mistake. When something goes wrong, we are good at protecting each other and focusing on system errors, not personal ones. [Norway, participant 2]

Limitations and Future Implications

Our study has several limitations. Qualitative research restricts the use of formal statistical analysis as broad-ranging emotions reduce its reproducibility. These challenges were amplified by complex sociopolitical–technological variations. Generalizability is limited owing to the small sample size and single-center analysis in each country. Therefore, we can only extrapolate speculations to explain the results of this study. No definitive conclusions can be made regarding EHR user satisfaction between the 2 countries. Although this study specifically recruited EM physicians, future research may benefit from expanding to other specialties across multiple facilities. Despite the semistructured interviews having reproducibility limitations, this method was necessary to understand the health care infrastructure and nuances of daily practice within each location. New questions emerged as more information was gained. Although this approach creates inconsistencies, it permits flexibility that is otherwise impossible to achieve using alternative qualitative methods such as surveys. These humanistic interactions are both a strength and weakness of semistructured interviews. Objective metrics regarding usability and satisfaction are difficult to produce with countless independent variables. Nevertheless, this comparison provides rich insight.

Numerous potential factors that may contribute to poor EHR user experiences were identified during the first phase of data collection (American interviews). Much of this occurred without prompts from the interviewer (GG). If these factors did not come up organically in Norwegian physician interviews, the interviewer asked targeted questions pertaining to these topics with the intention of identifying similarities or differences. Although this does not alter the United States’ findings, it may artificially inflate Norwegian results regarding perceived EHR disadvantages.

A study by Tutt et al [14] described factors that may enhance EHR experiences and suggested that policy makers, software developers, HIT vendors, payers, health administrators, and users alike may be capable of contributing to collective improvements. They also identified administrative tasks that add to documentation burden, including extensive order entries, billing regulations, coding standards, quality improvement reporting, and system security [14]. As Colicchio et al [40] noted, it is important to consider that national EHRs may not provide the desired insight for future informatics research, as local configurations are customizable even when supplied by the same vendor. After the Helseplatformen is implemented in Norway, prospective longitudinal studies measuring similar outcomes may produce additional meaningful information. This novel investigation suggests a framework for theoretical EHR optimization on a global scale. Although the results of this study are not entirely generalizable, it provides a foothold for future research and may stimulate innovative HIT advancements. Additional studies that compare international experiences while considering social and political differences are needed to identify the components that most significantly influence user satisfaction.

Conclusions

This qualitative study explores factors that influence EHR user satisfaction among practicing EM physicians in 2 countries. All the participants believed that this technology has increased their workload while simultaneously acknowledging their heavy reliance on it. They agreed that EHRs are here to stay. The results show that both American and Norwegian physicians experience frustration with EHR, but overall, the United States cohort had significantly more complaints. Participant-driven conversations revealed that each country had moderately differing sources of frustration. Norwegian complaints revolved around intrinsic technical issues. Strategies to mitigate these problems are currently underway as evidenced by the En Inbygger–En Journal and Helseplatformen initiative. Americans harshly criticized the business of medicine that they felt was manifested in every facet of HIT implementation. These findings enhance the theory that policies and administration may influence usability to a greater degree than technology itself [9,14,24,26,95].

Use of in-depth, semistructured interviews permitted a deeper understanding of both health care systems. This knowledge was subsequently integrated throughout data analysis and interpretation. The development and use of EHRs is influenced by lawmakers, payers, companies, and regulatory entities. Decisions made by those who are not primary users have a profound impact on the practice of those who use this technology daily. Both countries in this study are currently undergoing significant changes. Norway is poised to make a complete national overhaul of their EHR, and the United States is struggling to reform a vast, expensive, and inefficient health system. If HIT is to be optimized on a global scale, the elements highlighted in this study should be considered when establishing policy, strategy, and vision for the future.

Acknowledgments

The authors would like to thank the following individuals for helping them with participant recruitment and overall project support: Bradley Barth, MD; Peyman Usefi, MD; Marius Aagaard, MD; and Morgan Denton. The authors are grateful to all the participating physicians at the University of Kansas (KU) Health System and Akershus Universitetssykehus. They would also like to acknowledge Hanah Valeur, MD and Eirik Nikolai Arnesen, MD for providing insight into the Norwegian Medical Association and the structure...
of the Norwegian health care system. This research was supported by the Clendening and King Summer Fellowship through the KU School of Medicine–Department of History and Philosophy of Medicine. The processing charges related to the publication of this paper were supported by the KU One University Open Access Author Fund sponsored jointly by the KU Provost, KU Vice Chancellor for Research and Graduate Studies, and the KU Medical Center Vice Chancellor for Research and managed jointly by the Libraries at the Medical Center and KU-Lawrence.

Conflicts of Interest
None declared.

References


41. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. P T 2017 Sep;42(9):572-575 [Full text] [Medline: 28890644]


Abbreviations

ED: emergency department
EHR: electronic health record
EM: emergency medicine
HIE: health information exchange
HIT: health information technology
KU: University of Kansas
KUMC: University of Kansas Medical Center
NMA: Norwegian Medical Association
VA: US Department of Veterans Affairs

©Gracie Garcia, Christopher Crenner. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 07.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Combining Farmers’ Preferences With Evidence-Based Strategies to Prevent and Lower Farmers’ Distress: Co-design and Acceptability Testing of ifarmwell

Kate M Gunn1,2, BPsych (Hons), MPsych (Clin), PhD; Gemma Skaczkowski1, BHSc (Hons), PhD; James Dollman3, BSc, MSocSci, DipEd, PhD; Andrew D Vincent2, BSc (Hons), PhD; Camille E Short4, BPsych (Hons), PhD; Susan Brumby5,6, RN, RM, DipFarm, GDipWSt, MMG, MHlthMgmt, PhD; Alison Barrett1, MRes, BHlthSC, BIntSt; Nathan Harrison1, BPsych (Hons); Deborah Turnbull2,7, BA (Hons), MPsych (Clin), PhD

1Department of Rural Health, Allied Health and Human Performance, University of South Australia, Adelaide, Australia
2Freemason’s Centre for Male Health and Wellbeing, The University of Adelaide, Adelaide, Australia
3Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Performance, University of South Australia, Adelaide, Australia
4Melbourne Centre for Behaviour Change, Melbourne School of Psychological Sciences and Melbourne School of Health Sciences, University of Melbourne, Melbourne, Australia
5National Centre for Farmer Health, Western District Health Service, Hamilton, Australia
6School of Medicine, Deakin University, Melbourne, Australia
7School of Psychology, The University of Adelaide, Adelaide, Australia

Corresponding Author:
Kate M Gunn, BPsych (Hons), MPsych (Clin), PhD
Department of Rural Health
Allied Health and Human Performance
University of South Australia
City East Campus BJ2-03, North Terrace
Adelaide, 5001
Australia
Phone: 61 417852537
Email: kate.gunn@unisa.edu.au

Abstract

Background: Farming is physically and psychologically hazardous. Farmers face many barriers to help seeking from traditional physical and mental health services; however, improved internet access now provides promising avenues for offering support.

Objective: This study aims to co-design with farmers the content and functionality of a website that helps them adopt transferable coping strategies and test its acceptability in the broader farming population.

Methods: Research evidence and expert opinions were synthesized to inform key design principles. A total of 18 farmers detailed what they would like from this type of website. Intervention logic and relevant evidence-based strategies were mapped. Website content was drafted and reviewed by 2 independent mental health professionals. A total of 9 farmers provided detailed qualitative feedback on the face validity of the draft content. Subsequently, 9 farmers provided feedback on the website prototype. Following amendments and internal prototype testing and optimization, prototype usability (ie, completion rate) was examined with 157 registered website users who were (105/157, 66.9%) female, aged 21-73 years; 95.5% (149/156) residing in inner regional to very remote Australia, and 68.2% (107/157) “sheep, cattle and/or grain farmers.” Acceptability was examined with a subset of 114 users who rated at least module 1. Interviews with 108 farmers who did not complete all 5 modules helped determine why, and detailed interviews were conducted with 18 purposively sampled users. Updates were then made according to adaptive trial design methodology.

Results: This systematic co-design process resulted in a web-based resource based on acceptance and commitment therapy and designed to overcome barriers to engagement with traditional mental health and well-being strategies—ifarmwell. It was considered an accessible and confidential source of practical and relevant farmer-focused self-help strategies. These strategies were delivered via 5 interactive modules that include written, drawn, and audio- and video-based psychoeducation and exercises, as well as farming-related jokes, metaphors, examples, and imagery. Module 1 included distress screening and information on how to speak...
to general practitioners about mental health–related concerns (including a personalized conversation script). Modules were completed fortnightly. SMS text messages offered personalized support and reminders. Qualitative interviews and star ratings demonstrated high module acceptability (average 4.06/5 rating) and suggested that additional reminders, higher quality audio recordings, and shorter modules would be useful. Approximately 37.1% (52/140) of users who started module 1 completed all modules, with too busy or not got to it yet being the main reason for non-completion, and previous module acceptability not predicting subsequent module completion.

**Conclusions:** Sequential integration of research evidence, expert knowledge, and farmers’ preferences in the co-design process allowed for the development of a self-help intervention that focused on important intervention targets and was acceptable to this difficult-to-engage group.

**Trial Registration:** Australian New Zealand Clinical Trials Registry ACTRN12617000506392; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=372526

(MJM Hum Factors 2022;9(1):e27631) doi:10.2196/27631

**KEYWORDS**

farm; agriculture; rural; drought; mental health; stress; coping; online intervention; acceptance and commitment therapy

**Introduction**

**Background**

Farming is an occupation that involves numerous physical and psychological hazards. In recent years, Australian farmers have faced increased exposure to natural disasters, particularly prolonged droughts, fires and floods [1]. Farmers often both live and work on their farms, with family members across multiple generations being involved, consequently blurring the line between work, home, and family roles, which adds to their stress [2-5]. Financial pressure, loss of control, and uncertainty about the future are also associated with environmental stressors and are thought to significantly increase the risk of farmers experiencing mental health problems [3,6,7]. The inability to control these stressors and the sense of hopelessness and entrapment they can engender are thought to be potential risk factors for rural male suicide [8]. Indeed, studies have found a significantly higher incidence of suicide among rural and remote populations compared with metropolitan populations [9,10] and between agricultural workers compared with other employed rural people [11,12].

At the same time, farmers are known to face numerous barriers to help seeking from traditional physical and mental health services. These barriers are structural, such as the limited availability of medical and psychological professionals [13], and attitudinal [14,15]. For generations, Australian farmers have been characterized as being independent, stoic, and skilled at solving practical problems [2,16]. However, in the context of help seeking for the management of psychological distress, traits such as stoicism, independence, and a strong desire to keep personal matters private, may in fact be maladaptive [17]. Recent Australian research has found that farmers were half as likely to have sought help from a general practitioner (GP) or mental health professional in the previous 6 months compared with other employed rural people [13].

Fortunately, the National Broadband Network has now been rolled out in Australia, increasing rural access to internet sites and services [18]. A recent survey of 2000 businesses within the Australian agricultural sector found that up to 95% now have access to the internet [19], and the use of the internet to access health services is known to be increasing in the rural population [20,21].

The delivery of evidence-based interventions on the web offers opportunities to overcome some traditional barriers to help seeking faced by these populations. There is emerging evidence that computerized cognitive behaviour therapy (CCBT) interventions are acceptable in rural communities [22], and an unpublished example of a CCBT intervention designed to address anxiety, depression, and social functioning in Scottish farmers is Living Life to the Full (although it reported limited success) [23]. Given farmers’ numerous barriers to help seeking and the strong perception within the industry that outsiders (including health professionals [24]) fail to understand their needs and way of life, the development of such interventions needs to be done carefully. Consumer involvement in intervention design ensures that interventions are relevant, usable, and culturally appropriate for the target audience [25,26], which in turn can improve intervention success [27].

**Objective**

The purpose of this paper is to describe the co-design of content and functionality of a website that aims to help farmers adopt transferable coping strategies that are likely to help them effectively cope with stress. The second purpose of this research is to test the acceptability and feasibility of this website in a broader Australian farming population. The development of this website involved the sequential integration of research evidence, expert knowledge, and farmers’ preferences. Methodological guidance and examples such as the studies by O’Brien et al [28] and Short et al [29] and the work outlined in this paper, provide a transparent account of intervention co-design and development upon which other clinicians and researchers can build.

**Methods**

**Overview**

Ethics approval for this project was granted by the University of South Australia human research ethics committee (application ID 0000035637). A 9-stage co-design process that included the sequential validation and optimization of evidence and expert
opinion with farmers’ wants and preferences was used in a process similar to that described by Easton et al [30]. Each stage resulted in outputs (described in the Results section) that were used to inform the next stage of development. Figure 1 summarizes these stages.

**Figure 1. Development process for ifarmwell website.**

Stage 1: Evidence synthesis to inform key design principles

Stage 2: Finding out what Australian farmers want from a web-based well-being resource: A qualitative study (N=18)

Stage 3: Translating design principles and farmers’ preferences into intervention logic and draft website content

Stage 4: Checking the clinical accuracy, safety and relevance of website content with independent mental health professionals

Stage 5: Testing the face validity of the draft website with Australian farmers: Website content review (N=9)

Stage 6: Intervention build and internal testing

Stage 7: Retesting face validity with Australian farmers: Draft prototype review and advice on launch (N=9)

Stage 8a: Usability and acceptability testing of the prototype (quantitative data): Module completion (N=157) and satisfaction ratings (n=114)

Stage 8b: Usability and acceptability testing of prototype (qualitative data): Interviews with purposively sampled website users (N=18) and website users who did not complete all five modules (N=103)

Stage 9: Iterative design changes

In total, four key methodological approaches informed these stages: (1) synthesis of evidence from prior research to understand the problem and possible solutions (stages 1 and 2); (2) intervention mapping techniques to chart the logic of the intervention (including key acceptance and commitment therapy [ACT] processes or performance objectives, determinants of change, relevant behavior change strategies, and persuasive system design elements; stage 3) [31-33]; (3) a person-based approach via the involvement of farmers as co-designers [34,35] (stages 2, 5, and 7); and (4) iterative updating based on user feedback that allows for ongoing improvements to be made to the website (stages 8 and 9), which is informed by adaptive trial design methodology [36].

All farmers who participated in the research were adults who owned or played an active role in the operation of a farming or pastoral enterprise in Australia (or the spouse of someone who did), were fluent in English, had access to the internet, and had access to a mobile phone with reliable connection or reception at least once per week. The following 9-stage iterative process was conducted over a 3-year period.

**Stage 1: Evidence Synthesis to Inform Key Design Principles**

Key learnings from published works [6,14,37-40], our own unpublished work, and views from relevant experts across the agricultural, financial, and mental health fields were summarized by the research team. The research team was well-placed to prioritize learnings, given their extensive knowledge of agriculture (KMG, SB, JD, and AB), behavior change interventions (DT, CES, and KMG), web-based interventions (KG, CES, and SB), and rural health (KG, SB, JD, AB, and NH) and mental health (KMG and DT).

**Stage 2: Finding Out What Australian Farmers Want From a Web-Based Well-being Resource—A Qualitative Study**

**Participants**

A total of 11 male (11/18, 61%) and 7 female farmers (7/18, 39%), who met the above criteria, participated in the interviews. They had a median age of 45.5 years and were all from grain, sheep, and/or cattle farms across 4 Australian states.

**Procedures**

As described in detail elsewhere [41], participants were recruited via articles in print, radio and web-based media, advertising via relevant rural organizations, and personal and professional contacts of the research team. Telephone interviews were used to explore the farmers’ current internet use practices and preferences for websites designed to promote their mental health and well-being. Thematic analysis was used to analyze the verbatim interview transcripts [42]. Data were arranged under each theme in a Microsoft Excel spreadsheet using a framework approach. The data were checked for any evidence of themes that contradicted the key design principles identified in stage 1.

**Stage 3: Translating Design Principles and Farmers’ Preferences Into the Intervention Logic and Draft Website Content**

The logic of the intervention was systematically developed by KG to ensure that important intervention targets (identified in
stage 1 and explained further in the *Results* section) were addressed and that the effectiveness of the targets could be systematically assessed later. This included mapping the module content to the core ACT processes (acceptance, cognitive defusion, being present, self as context, values, and committed action [43]). It also included ensuring that relevant behavior change techniques [31] (outlined in the *Results* section) were included throughout to help address each of the behavioral determinants (ie, knowledge, skills, emotion, action planning, beliefs about capabilities, beliefs about consequences, motivation and goals, and memory, attention, or decision-making processes) thought to influence whether a user would successfully adopt the core ACT processes. The selection of these behavior change techniques was based on what has been previously shown to effectively address relevant behavioral determinants [31]. Although some overlap with behavior change techniques and persuasive system design elements is acknowledged, persuasive system design elements (as defined by Kelders et al [33] and outlined in the *Results* section) were also built into the intervention logic to help maximize user engagement and limit dropout.

The text, video, and audio content contained within each website module were then drafted by KG by integrating the key design principles from stage 1, farmers’ preferences established in stage 2, and the intervention logic identified in stage 3. Her first-hand experience of using ACT in her role as a clinical psychologist, living on a farm in a farming family, developing self-help mental health materials for rural populations, and formal training in intervention mapping, assisted with this process. The general principles of adult learning [44] were also considered.

**Stage 4: Checking the Clinical Accuracy and Safety of Website Content With Independent Mental Health Professionals**

**Participants**

A male social worker with a long history of supporting drought-affected farmers and knowledge of and experience using ACT clinically and a female clinical psychologist highly experienced with clinical and forensic mental health populations and in the use of ACT, participated in this stage of testing.

**Procedures**

Independent feedback on the clinical accuracy, safety, and relevance of website content was provided on all website content using tracked changes in a Microsoft Word processing document. Suggestions were then incorporated where feasible (ie, would not make the modules too long) to enhance clinical impact.

**Stage 5: Testing the Face Validity of the Draft Website With Australian Farmers—Website Content Review**

**Participants**

A total of 9 farmers (4/9, 44% men and 5/9, 56% women), who met the criteria outlined above and had participated in stage 2, took part in this stage of the research (herein referred to as co-designers). They ranged in age from 34 to 62 years and were from grain, sheep or cattle properties in the states of South Australia (7/9, 78%) and Western Australia (2/9, 22%).

**Procedures**

A copy of the draft website content was sent to the co-designers via post or email. Participants were also asked to comment specifically on several logo and design options (colors, fonts, background images, and layouts) provided as PDF files. Interviews were then conducted over the phone (or, in one case, in person) to gather feedback, with a focus on language, relevance, and face validity.

**Analysis**

Where possible, key recommendations for improvement were compiled, and edits were made to the working draft document following the completion of each interview.

**Stage 6: Intervention Build and Internal Testing**

The purpose of this stage was to produce a working intervention prototype. The research team supplied the website content and design documents developed in earlier steps to a web developer and then worked in close collaboration with them to ensure that lessons from previous stages were integrated into the website and technical glitches were addressed. The prototype was made public in February 2018.

**Stage 7: Retesting the Face Validity of the Draft Website With Australian Farmers—Website Prototype Review and Advice on Launch**

**Participants**

A total of 4 farmers (co-designers; 1/4, 25% male and 3/4, 75% female) provided detailed feedback on the website prototype. They were aged 24, 40, 61, and 62 years and were from grain, sheep, and/or cattle properties in South Australia (2/4, 50%), Western Australia (1/4, 25%), and New South Wales (1/4, 25%). A further 5 farmer co-designers (all men) provided feedback specifically on the website launch. They were aged 34, 44, 47, 53, and 55 years and were from sheep or cattle properties (1/5, 20%) or grain, sheep and/or cattle properties (4/5, 80%) in South Australia.

**Procedures**

Co-designers were sent a link to the website prototype along with broad instructions to work through the website and provide email or phone comments on any aspects they thought required changing.

**Analysis**

Key recommendations from participant comments were compiled and implemented where possible.

**Stage 8a: Usability and Acceptability Testing of the Prototype (Quantitative Data)**

The trial was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617000506392) on April 3, 2017.
Participants
Usability testing of the prototype was conducted by 157 farmers who registered during the study period and met the criteria outlined above. Acceptability testing was undertaken with a subset of 114 users who provided a rating out of 5 for at least module 1. Their demographic characteristics are shown in Table 1.

Table 1. Demographics for all eligible registered users and those users who provided acceptability ratings for at least one module (stage 8a).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All registered users (N=157)</th>
<th>Users who provided acceptability ratings (N=114)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong> (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD)</td>
<td>45.55 (12.17)</td>
<td>45.46 (12.65)</td>
</tr>
<tr>
<td>Values, median (range)</td>
<td>46 (21-73)</td>
<td>46 (21-73)</td>
</tr>
<tr>
<td><strong>Gender</strong> (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>105 (66.9)</td>
<td>79 (69.3)</td>
</tr>
<tr>
<td>Male</td>
<td>52 (33.1)</td>
<td>35 (30.7)</td>
</tr>
<tr>
<td><strong>Remoteness of residence</strong> (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major cities of Australia</td>
<td>7 (4.5)</td>
<td>6 (5.3)</td>
</tr>
<tr>
<td>Inner regional Australia</td>
<td>66 (42.3)</td>
<td>46 (40.7)</td>
</tr>
<tr>
<td>Outer regional Australia</td>
<td>59 (37.8)</td>
<td>43 (38.1)</td>
</tr>
<tr>
<td>Remote Australia</td>
<td>16 (10.3)</td>
<td>13 (11.5)</td>
</tr>
<tr>
<td>Very remote Australia</td>
<td>8 (5.1)</td>
<td>5 (4.4)</td>
</tr>
<tr>
<td><strong>Farm type</strong> (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dairy</td>
<td>19 (12.1)</td>
<td>15 (13.2)</td>
</tr>
<tr>
<td>Grain, sheep and/or cattle</td>
<td>63 (40.1)</td>
<td>43 (37.7)</td>
</tr>
<tr>
<td>Horticulture, market garden, or fruit</td>
<td>14 (8.9)</td>
<td>9 (7.9)</td>
</tr>
<tr>
<td>Poultry</td>
<td>3 (1.9)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Sheep and/or cattle</td>
<td>44 (28)</td>
<td>37 (32.5)</td>
</tr>
<tr>
<td>Viticulture</td>
<td>1 (0.6)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Other</td>
<td>13 (8.3)</td>
<td>7 (6.1)</td>
</tr>
<tr>
<td><strong>Education level (highest qualification)</strong> (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>17 (10.8)</td>
<td>13 (11.4)</td>
</tr>
<tr>
<td>University degree or diploma</td>
<td>70 (44.6)</td>
<td>49 (43)</td>
</tr>
<tr>
<td>Trade certificate</td>
<td>43 (27.4)</td>
<td>34 (29.8)</td>
</tr>
<tr>
<td>Finished high school</td>
<td>25 (15.9)</td>
<td>17 (14.9)</td>
</tr>
<tr>
<td>Finished primary school</td>
<td>2 (1.3)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Hours per week spent using the internet, mean (SD)</td>
<td>16.42 (10.47)</td>
<td>16.07 (10.23)</td>
</tr>
</tbody>
</table>

*\(n=156\) and \(n=113\) because of missing data.

*\(n=155\) because of missing data.

Procedures
Consent for participation was established when users registered with the website. Data were collected from all users who registered between February and October 2018 inclusive.

Analysis
Analyses were conducted using SPSS Statistics for Windows (version 26; IBM Corp) [45]. Usability and acceptability were captured in several ways.

Star Ratings (Out of 5) by Each User at the Completion of a Module
At the end of each module, users were asked to rate that module on a scale ranging from 1 to 5 stars, where 1=unhelpful, 2=neutral, 3=satisfactory, 4=helpful, and 5=very helpful. The star rating out of 5 was used as it allowed for the multifaceted nature of acceptability to be captured [46] and because of the familiarity and briefness of this approach [47]. Acceptability ratings were examined for modules completed between February and October 2018. Ratings of acceptability for each module were estimated through a linear mixed model with maximum likelihood estimation, and the module number was entered as
a fixed effect with 5 levels and a random intercept per participant. Baseline age, gender, education, farm type, remoteness, hours of internet use, psychological distress, and stress were also entered as fixed factors. The average acceptability rating for each user was calculated from the star ratings of all modules that a user completed.

**Module Completion Rate**

Data on module completion were captured beyond the February to October 2018 time frame (up to February 2020) to capture participants’ full record of participation (even if this was post-October 2018).

**Association of Module Completion and Acceptability With Participant Demographics, Recent Exposure to Stressors and Distress Levels**

During the registration process, demographics (gender, age, education level, and farm type), distress (Kessler Psychological Distress Scale [48]), and a single-item measure of exposure to stressors were completed. For the latter, users were asked to think of the most stressful situation they had encountered during the past month and rate how stressful they found this situation on a scale of 1 to 10 [40]. Residential postcodes were used to calculate remoteness using the Accessibility and Remoteness Index of Australia from the Australian Bureau of Statistics [49]. Owing to small numbers, the categories finished high school and finished primary school were combined for analysis. Similarly, poultry farming and viticulture were grouped with other farm type.

The association between demographics, stress exposure, distress, and module acceptability was examined using the mixed model described above. A series of univariable and multivariable linear regressions examined the relationship between the number of modules completed and demographic or distress and stress variables. Finally, Pearson correlations were used to examine the association between module completion and an individual’s average acceptability rating and the rating of the last module they completed.

**Stage 8h: Usability and Acceptability Testing of the Prototype (Qualitative Data)**

**Participants and Procedures**

**Brief Phone Calls With Users Who Did Not Complete All 5 Modules (to Find Out Why)**

A total of 108 website users who had not continued with the next module within 5 weeks of completing the previous module were followed up with 2 phone calls, 1 email, and 1 additional attempt via email or phone approximately 1 month after that. Successful follow-ups were used to determine the reasons for not continuing with the modules so that we could find ways to enhance the website and aid engagement. Verbatim notes were taken during the phone calls along with email responses, which were manually analyzed by AB and KG using conventional content analysis [50] and a Microsoft Excel spreadsheet. Categories were derived from the data and reworked until all the data could be accounted for. Discrepancies between the coders were rare but were worked through until full agreement was reached.

**Detailed Phone Interviews With Purposively Sampled Group of Users**

A total of 18 farmers (7/18, 39% men and 11/18, 61% women) who had used the website were purposively selected from website users to gain a variety of impressions (based upon state, farm type, average module acceptability score, gender, and age) and invited via email to take part in a telephone interview to share their experiences. Farmers ranged in age from 23 to 71 years and were from dairy (1/18, 6%), horticulture (2/18, 11%), viticulture (1/18, 6%), sheep and/or cattle properties (7/18, 39%), and grain, sheep and/or cattle properties (6/18, 33%) in Victoria (6/18, 33%), New South Wales (4/18, 22%), South Australia (2/18, 11%), Tasmania (2/18, 11%), Western Australia (2/18, 11%), and Queensland (1/18, 6%). Interviews were audio recorded, transcribed verbatim, and analyzed by AB and KG using thematic analysis [42], with data arranged in a Microsoft Excel spreadsheet using a framework approach, and any discrepancies in coding discussed and reworked until full agreement was reached.

**Stage 9: Iterative Design Changes**

Following the acceptability assessment of the prototype outlined above, the website was adapted to improve user experience. This aligns with the adaptive trial design methodology [36] and the person-based approach to intervention design of Yardley [34] by continuing to incorporate user feedback after live testing of the intervention.

**Results**

**Stage 1: Evidence Synthesis to Inform Key Design Principles**

A summary of our evidence synthesis and the key overarching design principles identified from this are shown in Table 2. In brief, farmers face many barriers to accessing traditional face-to-face mental health services, including a lack of service availability, cost, time, and concerns about confidentiality. They also perceive that outsiders (including health professionals) often do not understand the issues they face. The types of challenges that cause farmers the most stress are those that are beyond their control, and these are the things they feel least equipped to cope effectively with. However, acceptance has been shown to be an adaptive coping strategy for farmers in this context [40]. Together, these factors suggest that a new web-based mental health and well-being resource could help overcome existing barriers to engagement by being an accessible, confidential source of farmer-focused, practical self-help strategies based on ACT [51] if co-designed with farmers.

https://humanfactors.jmir.org/2022/1/e27631

---

**Stage 8b: Usability and Acceptability Testing of the Prototype (Qualitative Data)**

**Participants and Procedures**

**Brief Phone Calls With Users Who Did Not Complete All 5 Modules (to Find Out Why)**

A total of 108 website users who had not continued with the next module within 5 weeks of completing the previous module were followed up with 2 phone calls, 1 email, and 1 additional attempt via email or phone approximately 1 month after that. Successful follow-ups were used to determine the reasons for not continuing with the modules so that we could find ways to enhance the website and aid engagement. Verbatim notes were taken during the phone calls along with email responses, which were manually analyzed by AB and KG using conventional content analysis [50] and a Microsoft Excel spreadsheet. Categories were derived from the data and reworked until all the data could be accounted for. Discrepancies between the coders were rare but were worked through until full agreement was reached.

**Detailed Phone Interviews With Purposively Sampled Group of Users**

A total of 18 farmers (7/18, 39% men and 11/18, 61% women) who had used the website were purposively selected from website users to gain a variety of impressions (based upon state, farm type, average module acceptability score, gender, and age) and invited via email to take part in a telephone interview to share their experiences. Farmers ranged in age from 23 to 71 years and were from dairy (1/18, 6%), horticulture (2/18, 11%), viticulture (1/18, 6%), sheep and/or cattle properties (7/18, 39%), and grain, sheep and/or cattle properties (6/18, 33%) in Victoria (6/18, 33%), New South Wales (4/18, 22%), South Australia (2/18, 11%), Tasmania (2/18, 11%), Western Australia (2/18, 11%), and Queensland (1/18, 6%). Interviews were audio recorded, transcribed verbatim, and analyzed by AB and KG using thematic analysis [42], with data arranged in a Microsoft Excel spreadsheet using a framework approach, and any discrepancies in coding discussed and reworked until full agreement was reached.

**Stage 9: Iterative Design Changes**

Following the acceptability assessment of the prototype outlined above, the website was adapted to improve user experience. This aligns with the adaptive trial design methodology [36] and the person-based approach to intervention design of Yardley [34] by continuing to incorporate user feedback after live testing of the intervention.

**Results**

**Stage 1: Evidence Synthesis to Inform Key Design Principles**

A summary of our evidence synthesis and the key overarching design principles identified from this are shown in Table 2. In brief, farmers face many barriers to accessing traditional face-to-face mental health services, including a lack of service availability, cost, time, and concerns about confidentiality. They also perceive that outsiders (including health professionals) often do not understand the issues they face. The types of challenges that cause farmers the most stress are those that are beyond their control, and these are the things they feel least equipped to cope effectively with. However, acceptance has been shown to be an adaptive coping strategy for farmers in this context [40]. Together, these factors suggest that a new web-based mental health and well-being resource could help overcome existing barriers to engagement by being an accessible, confidential source of farmer-focused, practical self-help strategies based on ACT [51] if co-designed with farmers.
Table 2. Design principles resulting from the evidence synthesis.

<table>
<thead>
<tr>
<th>Evidence synthesis</th>
<th>Resulting design principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barriers to accessing face-to-face mental health and well-being services in rural</td>
<td>Web-based resources may help to address barriers to the access and availability of services.</td>
</tr>
<tr>
<td>areas include cost, time, stigma, a lack of anonymity in country towns, a general</td>
<td></td>
</tr>
<tr>
<td>lack of understanding of mental health issues, and the lack of availability of</td>
<td></td>
</tr>
<tr>
<td>services [5,16,52-58].</td>
<td></td>
</tr>
<tr>
<td>Barriers to help seeking for mental health issues among farmers include the desire</td>
<td>Self-help resources align with farmers’ desire for control, self-reliance, and anonymy.</td>
</tr>
<tr>
<td>for control, self-reliance, tendency to minimize the problem, and resignation [14,</td>
<td></td>
</tr>
<tr>
<td>37]. Farmers prefer anonymous self-help books or internet resources [59].</td>
<td></td>
</tr>
<tr>
<td>Farmers are often isolated and perceive a lack of understanding of rural issues</td>
<td>Having a clear farming focus and co-designing alongside farmers is needed to ensure</td>
</tr>
<tr>
<td>from outsiders [6,38].</td>
<td>relevance and acceptability.</td>
</tr>
<tr>
<td>Many farmers report difficulty understanding health care professionals [14] and</td>
<td></td>
</tr>
<tr>
<td>that health care professionals do not understand them and their way of life [6,53].</td>
<td></td>
</tr>
<tr>
<td>However, there is a high level of community trust within rural Australia [39],</td>
<td></td>
</tr>
<tr>
<td>suggesting that a resource designed by farmers and for farmers may be considered</td>
<td></td>
</tr>
<tr>
<td>credible.</td>
<td></td>
</tr>
<tr>
<td>Managing uncertainty is a key challenge resulting from drought and a stressor that</td>
<td>Uncertainty about the future is a key stressor that farmers need help with managing.</td>
</tr>
<tr>
<td>many farmers do not feel equipped to manage [6]. They are generally already good</td>
<td>An interactive, engaging resource is needed.</td>
</tr>
<tr>
<td>at solving problems, so they are less likely to benefit from assistance with that.</td>
<td>Acceptance is an effective coping strategy for farmers in this context.</td>
</tr>
<tr>
<td>Information provision and educational resources alone are not enough to change key</td>
<td>ACT may be an appropriate therapeutic model for this context.</td>
</tr>
<tr>
<td>behaviors and thought processes [60]. Evidence-based behavior change techniques</td>
<td></td>
</tr>
<tr>
<td>(eg, modeling, self-monitoring, and goal setting) should be built into web-based</td>
<td>Issues relating to web-based intervention adherence need to be addressed.</td>
</tr>
<tr>
<td>interventions to maximize the effect [33,61].</td>
<td></td>
</tr>
<tr>
<td>Farmers who adopt acceptance as a coping strategy and do not engage in behavioral</td>
<td></td>
</tr>
<tr>
<td>disengagement (giving up) are less likely to experience distress when faced with</td>
<td></td>
</tr>
<tr>
<td>significant stressors during drought [40].</td>
<td></td>
</tr>
<tr>
<td>ACT(^a) is a transdiagnostic, evidence-based psychotherapeutic approach that can</td>
<td></td>
</tr>
<tr>
<td>foster acceptance and committed action (opposite of giving up) and improve well-</td>
<td></td>
</tr>
<tr>
<td>being in a nonpathologizing way [62]. ACT may be used to address a range of</td>
<td></td>
</tr>
<tr>
<td>psychological disorders and promote general well-being in nonclinical samples [62-64],</td>
<td></td>
</tr>
<tr>
<td>including via web-based interventions [64,65]. It is particularly suited to</td>
<td></td>
</tr>
<tr>
<td>contexts where the stressor must be accepted or cannot be fixed [66].</td>
<td></td>
</tr>
<tr>
<td>Strategies to improve intervention adherence and effectiveness must also be included</td>
<td></td>
</tr>
<tr>
<td>(eg, tunneling, personalization, and reminders) [33,67-69].</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ACT: acceptance and commitment therapy.

Stage 2: Finding Out What Australian Farmers Want From Web-Based Well-being Resources—a Qualitative Study

As reported elsewhere [41], farmers said that they would like a web-based resource that is easy to navigate and compatible with multiple devices and internet connections, as well as their sporadic internet use around work schedules. They preferred a casual and friendly tone, minimal use of jargon, and the inclusion of humor, and they requested information on when and how to seek additional professional help. They also said that they wanted a resource that was authentic, that reflected their challenges and way of life, and that they could see the benefits from quickly. There was no evidence of themes that contradicted the key design principles identified in stage 1.

Stage 3: Translating Design Principles and Farmers’ Preferences Into the Intervention Logic and Draft Website Content

Overview

The resulting ifarmwell web-based intervention is a free, farmer-focused, password-protected self-help resource that contains 5 modules. Textbox 1 outlines the purpose of each module as explained to users, and Table 3 details the intervention logic and design, including the key content, targeted ACT processes, behavior change techniques, and persuasive system design elements contained within each module. The content is written for a low reading age (Gunning Fog score=5.8, easily understood by individuals aged 13-14 years) using friendly language with appropriate humor and farming-related metaphors and examples and fits with farmers’ ethos of independence and determination to help themselves. The intervention is nonpathologizing and focuses on improving well-being and preventing poor mental health rather than treating poor mental health or mental illness. The word mental health is avoided where possible on the website based on farmers’ advice about how best to engage their peers.
**Textbox 1.** ifarmwell module aims (as presented to users).

<table>
<thead>
<tr>
<th>Module 1: Taking stock of your current well-being and some practical strategies to get you started</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Confidentially discover how your current well-being compares with the well-being of other Australians</td>
</tr>
<tr>
<td>• Learn about additional support services that may be useful for you in addition to this web-based resource</td>
</tr>
<tr>
<td>• Provide some practical strategies tailored to specific challenges you may face</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module 2: Thoughts are like bullies—how to spend less time in your head</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Understand the power thoughts have over the way you feel</td>
</tr>
<tr>
<td>• Become more aware of the thoughts or stories your mind plays to you</td>
</tr>
<tr>
<td>• Learn how to look at your thoughts rather than from them</td>
</tr>
<tr>
<td>• Practice evaluating whether a particular thought is helpful to tune in to or not</td>
</tr>
<tr>
<td>• Start to learn how to let go of unhelpful thoughts and focus on things that make life better</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module 3: Doing what really matters—how to get the most out of life</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Work out what is important to you</td>
</tr>
<tr>
<td>• Identify areas of life in which it would be useful to put more energy</td>
</tr>
<tr>
<td>• Recognize areas of your life in which it might be useful to put less energy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module 4: Training your attention muscle and focusing on the here and now—a more pleasant, less exhausting place to be</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Become more aware of where your attention is and how this affects how you feel and behave</td>
</tr>
<tr>
<td>• Practice shifting attention to the here and now</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module 5: Putting it all together and moving forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Revisit strategies</td>
</tr>
<tr>
<td>• Plan out how to build these new strategies into day-to-day life</td>
</tr>
<tr>
<td>• Think about situations where familiar thoughts or stories may be triggered</td>
</tr>
<tr>
<td>• Plan how to respond to new challenges</td>
</tr>
</tbody>
</table>
Table 3. ifarmwell intervention logic and design.

<table>
<thead>
<tr>
<th>ACT&lt;sup&gt;a&lt;/sup&gt; processes</th>
<th>Behavior change techniques (targeting key behavioral determinants of adoption of ACT processes)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Persuasive system design elements (to aid engagement)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Content details</th>
</tr>
</thead>
</table>
| Module 1: taking stock of your current well-being and some practical strategies to get you started | No ACT processes targeted | • Self-monitoring  
• Persuasive communication  
• Information regarding outcomes  
• Personalized messages  
• Modeling or demonstration  
• Goal setting or homework | • Reduction  
• Tunneling  
• Tailoring  
• Personalization  
• Self-monitoring  
• Praise  
• Reminders  
• Suggestion  
• Similarity  
• Liking  
• Social learning  
• Normative influence | • Feedback from K10<sup>d</sup> (current levels of distress) and COPE<sup>e</sup> (current coping strategies)  
• Personalized script for discussion with GP<sup>f</sup> (if medium or high level of distress identified)  
• Video demonstration of farmer speaking to GP about mental health using a script  
• Psychoeducation tip sheets for 3 user-identified challenges  
• Basic self-care and helpful coping strategies (default)  
• Improving the quality of your sleep  
• Managing conflict with others  
• Improving the quality of your relationship  
• How to get your point across  
• Managing anger  
• Coping with grief and loss  
• Alcohol and drug use  
• Dealing with domestic violence  
• Adapting to new roles  
• What to do if you are feeling down or low  
• Coping after a natural disaster  
• Succession planning  
• Feeling trapped in an unhappy relationship  
• What to expect in upcoming modules (intro to ACT)  
• Homework planning or goal setting to implement tip sheet strategies |

Module 2: thoughts are like bullies—how to spend less time in your head

- Acceptance  
- Cognitive defusion  
- Being present  
- Self as context (being aware of your experiences without being attached to them)
- Personalized messages  
- Information regarding outcomes  
- Self-monitoring  
- Rewards or positive feedback (encouragement or reinforcement)  
- Problem-solving  
- Persuasive communication  
- Prompts, triggers, and cues  
- Rehearsal of relevant skills  
- Graded tasks  
- Goal setting or homework
- Reduction  
- Tunneling  
- Tailoring  
- Personalization  
- Self-monitoring  
- Praise  
- Reminders  
- Suggestion  
- Similarity  
- Liking  
- Social learning  
- Rehearsal
- Homework review or problem-solving obstacles  
- Feedback from Automatic Thoughts Questionnaire (identification of key challenging stories)  
- Exploration of existing strategies tried to manage challenging stories. Worked?  
- Pink sheep or elephants exercise; creative hopelessness  
- Video: piece-of-paper metaphor demonstration  
- Audio: notice thoughts while breathing (tool 1)  
- Examining whether particular thoughts are helpful to focus on or not (drag and drop task with feedback; tool 2)  
- Drafting thoughts in to just do it, plan a time, and let it go pens  
- “I’m having the thought that...” exercise (tool 3)  
- Giving stories a name exercise (tool 4)  
- Identifying thinking errors (tool 5)  
- Additional strategies to help you think differently about your thoughts (extra metaphors; tool 6)  
- Homework planning or goal setting to implement strategies
<table>
<thead>
<tr>
<th>ACT(^a) processes</th>
<th>Behavior change techniques (targeting key behavioral determinants of adoption of ACT processes)(^b)</th>
<th>Persuasive system design elements (to aid engagement)(^c)</th>
<th>Content details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 3: Doing what really matters—how to get the most out of life</td>
<td>- Values - Committed action</td>
<td>- Reduction - Tunneling - Tailoring - Personalization - Self-monitoring - Praise - Reminders - Suggestion - Similarity - Liking</td>
<td>- Homework review or problem-solving obstacles - Consideration of current influences on behavior - Valuing questionnaire and tailored feedback (removed, stage 9) - Values clarification (drag and drop task) - Reflection on current and future decision-making and interactions with others and considering values (tool 7) - Planning to live more consistently with top 10 values in next week and next 6 months (acknowledge what already doing, schedule time, and plan to overcome obstacles; tool 8) - Homework planning or goal setting to implement strategies</td>
</tr>
<tr>
<td>Module 4: training your attention muscle and focusing on the here and now—a more pleasant, less exhausting place to be</td>
<td>- Being present - Acceptance - Cognitive defusion - Self as context - Values - Committed action</td>
<td>- Reduction - Tunneling - Tailoring - Personalization - Self-monitoring - Praise - Reminders - Suggestion - Similarity - Liking - Rehearsal</td>
<td>- Homework review or problem-solving obstacles - Identifying existing activities fully present - Audio: here and now exercise (tool 9) - The basic (mindfulness) formula (tool 10) - Audio: 5 slow, deep breaths grounding technique (tool 11) - Audio: notice 3 things grounding technique (tool 12) - Paying attention to 1 thing at a time when doing everyday activities (tool 13) - Audio: letting go of difficult emotions (tool 14) - Homework planning or goal setting to implement strategies</td>
</tr>
<tr>
<td>Module 5: putting it all together and moving forward</td>
<td>- Acceptance - Cognitive defusion - Being present - Self as context - Values - Committed action</td>
<td>- Reduction - Tunneling - Tailoring - Personalization - Self-monitoring - Praise - Reminders - Suggestion - Similarity - Liking - Social learning - Rehearsal</td>
<td>- Homework review or problem-solving obstacles - Audio: leaves on a stream metaphor (tool 15) - Video: normalizes difficulty in mastering these strategies and encourages persistence - Summary of strategies (tool 16) - Audio: cows on a truck metaphor (tool 17) - Relapse prevention (warning signs): how to get yourself back on track and who you could turn to for extra help</td>
</tr>
</tbody>
</table>

\(^a\)ACT: acceptance and commitment therapy. 
\(^b\)As defined by Michie et al [31]. 
\(^c\)As defined by Kelders et al [33]. 
\(^d\)K10: Kessler Psychological Distress Scale. 
\(^e\)The COPE inventory [70]. 
\(^f\)GP: general practitioner.
The intervention was completed over 10 weeks, with each module taking approximately 30 minutes. Users could access the intervention at any time and on any device with an internet connection (e.g., laptop, desktop, tablet, and mobile phone). As shown in Figure 2, each module must be completed for the next module to be unlocked. This provided users with time to implement the strategies they learned from the previous module before moving to the next. This design was based on the literature showing that tunneled web-based interventions are less likely to overwhelm users and are better placed to personalize the intervention, leading to greater behavior change [67,68]. Figure 2 also indicates the frequency and type of SMS text messaging reminders sent to users throughout the intervention. Figure 2 shows the final design following changes made after the acceptability testing of the prototype (described in Stage 7).

Figure 2. Wireframe of the ifarmwell website.
**Personalization**

Tailored content was delivered throughout the intervention based on user responses and demographic variables. This included personalized imagery reflective of participants’ farming type, which has also been successfully used in farmer suicide stigma research [69]. In module 1, users were asked to complete the Kessler Psychological Distress Scale measure of distress and were provided with feedback about their current levels of distress, how these compared with others’ scores, and inform them if their scores suggested that they should seek professional face-to-face help. More specifically, based on their distress score, users were advised whether they were experiencing what was considered a low (10-19), mild (20-24), moderate (25-29), or severe (30-50) level of distress [71] and subsequently, whether it was recommended that they see their GP to discuss their well-being. Users were given the option of printing off the results of their web-based assessment and a script to guide a conversation about their mental health with their GP. Users were also presented with a short video showing someone else having that conversation with their GP. Finally, any severely distressed users (defined by cut-off) were contacted by a member of the research team by phone or email to encourage them to see a GP and remind them of helpline numbers.

At the end of each module, key tools were summarized, and users could choose to save them to their Toolbox if they found them useful. Modules 2 to 5 contained a homework review component, which asked users about the things they chose to focus on in the previous module and how much they had practiced them since. This section also asked users to reflect on whether anything got in the way or made doing this difficult and what they could do in the next week to overcome these difficulties.

At the end of the intervention, the Toolbox provided a summary of the user’s existing coping strategies, the stories the user’s mind often plays to them, their new preferred tools, and their top values (to guide future decision-making).

**Module Content**

Module content was transdiagnostic and useful for people experiencing a range of problems or conditions and for people simply wanting to improve their well-being or get more out of life. More specifically, module 1 was designed to take stock of users’ current well-being, suggest other sources of help if required, address basic self-care, and provide practical coping strategies that are targeted at users’ pressing, unique needs. This was based on a brief suite of questions used to identify the top 3 areas of need for each user. They were then presented with corresponding evidence-based tip sheets (eg, on sleep).

The remaining modules each focused on a particular ACT process. Module 2 addressed the power of thoughts and explained that avoidance or attempts to control difficult thoughts and feelings could be counterproductive. The module asked users to list the emotions and thoughts that they were struggling with, name the stories that they tell themselves, classify them as helpful to focus on, and identify errors in their thinking. Module 3 helped users clarify their values and find ways to live more consistently with them. Module 4 involved several mindfulness-based exercises (not labeled mindfulness-based upon farmers’ advice), designed to help users identify where their attention was and how this influenced them and practice shifting their attention to the here and now. Module 5 summarized the key strategies learned, examined possible triggers and warning signs to be aware of in the future, and reminded users of key sources of support.

**Stage 4: Checking the Clinical Accuracy and Safety of the Website Content With Independent Mental Health Professionals**

Mental health professionals provided guidance on the appropriateness and safety of the content and suggested minor changes. These included grammatical edits, alterations to simplify the language (eg, being clear about your values changed to knowing what matters to you), and adding a few more detailed explanations and metaphors to explain key concepts (eg, your mind as an ideas generator). Additional reflective questions were also suggested, for example, “what happened to the thought?” following an exercise to help let go of distressing thoughts. It was also recommended that additional text be added to help normalize the fact that one’s ability to focus and shift attention may vary from day to day.

**Stage 5: Testing the Face Validity of the Draft Website With Australian Farmers—Website Content Review**

Overall, participants felt that the module content was acceptable and relevant to farmers. Changes made to the content included repeating icons throughout the modules to guide the user, the inclusion of a summary of content at the beginning of each module, the inclusion of additional cartoons, and the removal of some references to stress, which farmers felt their peers would find off-putting (eg, under pressure rather than stressed). Additional methods for tailoring the content to farmers were also identified. For example, a co-designer suggested likening sorting out thoughts into different categories, to drafting sheep into different pens.

**Stage 6: Intervention Build and Internal Testing**

A web-based intervention prototype that could be tested by users was created, and wireframes to summarize the website’s structure were developed, as detailed in Figure 2. Internal testing by members of the research team resulted in comprehensive lists of hundreds of technical revisions that needed to be made by web developers to improve the user experience.

**Stage 7: Retesting the Face Validity of the Draft Website With Australian Farmers—Website Prototype Review and Advice on Launch**

This stage resulted in several changes to the look and feel of the website, such as a change of font color to improve readability and the inclusion of additional banner photographs featuring machinery and images of younger farmers to ensure broad appeal. Suggestions for improvement also included some website usability issues, such as the ease of saving and returning to a module later. Guidance was also provided on when would be a suitable time of year to advertise and launch the website (ie, not in January when many Australian grain farmers are on
holidays after busy harvests in the lead up to Christmas). The website was made public in February 2018.

**Stage 8a: Usability and Acceptability Testing of Prototype (Quantitative Data)**

**Module Completion Rate**

A total of 157 users (described in Table 4) registered on the website between February 2018 and October 2018 and were eligible to participate in the study. Of the 157 users, 17 (10.8%) users registered but did not start module 1. Table 4 shows the total number of people starting and completing each module.

The completion rates for modules 1 to 5 among those that commenced each module were 83.6% (117/140), 89% (81/91), 94% (68/72), 100% (58/58), and 100% (52/52), respectively. Approximately 35% (49/140) of the people who started module 1 did not start module 2 (dropout). The dropout rates for modules 2 to 4 were 21% (19/91), 19% (14/72), and 10% (6/58), respectively. Overall, 37.1% (52/140) of the people who started module 1 completed the entire intervention. The median time between starting module 1 and starting module 5 was 16 weeks (8 weeks was intended if users had 2 weeks break before commencing the next module), with a range of 8 to 76 weeks (15/52, 29% took 8-12 weeks; 28/52, 54% took 13-24 weeks; and 9/52, 17% took >24 weeks).

<table>
<thead>
<tr>
<th>Module</th>
<th>Started</th>
<th>Completed</th>
<th>Completion rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>140</td>
<td>117</td>
<td>83.6</td>
</tr>
<tr>
<td>Module 2</td>
<td>91</td>
<td>81</td>
<td>89</td>
</tr>
<tr>
<td>Module 3</td>
<td>72</td>
<td>68</td>
<td>94.4</td>
</tr>
<tr>
<td>Module 4</td>
<td>58</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Module 5</td>
<td>52</td>
<td>52</td>
<td>100</td>
</tr>
</tbody>
</table>

**Star Ratings (Out of 5) Provided by Each User at the Completion of a Module**

A total of 310 acceptability star ratings were submitted by 114 unique users (those who completed at least module 1 before October 2018). Of 114 users, the average rating across all modules on a 1- to 5-star rating scale was 4.06 (SD 0.99), with 17 (14.9%) people providing an average rating of 1 to 3, 43 (37.7%) people providing an average rating of >3 to 4, and 54 (47.4%) people providing an average rating of >4. The adjusted acceptability ratings for each module from the linear mixed model are shown in Table 5, and the mixed model is shown in Multimedia Appendix 1. There was a significant difference between module ratings; the module 3 acceptability rating was significantly lower than modules 1 (\( \beta =0.52, 95\%\ CI 0.28-0.77; P<.001\)), 2 (\( \beta = 0.58, 95\%\ CI 0.32-0.83; P<.001\)), 4 (\( \beta =0.49, 95\%\ CI 0.20-0.78; P=.001\)), and 5 (\( \beta =0.75, 95\%\ CI 0.44-1.08; P<.001\)).

<table>
<thead>
<tr>
<th>Module</th>
<th>Value, mean (SE; 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>4.01 (0.12; 3.77-4.25)</td>
</tr>
<tr>
<td>Module 2</td>
<td>4.06 (0.14; 3.79-4.34)</td>
</tr>
<tr>
<td>Module 3</td>
<td>3.49 (0.15; 3.19-3.78)</td>
</tr>
<tr>
<td>Module 4</td>
<td>3.98 (0.16; 3.67-4.29)</td>
</tr>
<tr>
<td>Module 5</td>
<td>4.25 (0.17; 3.91-4.59)</td>
</tr>
</tbody>
</table>

**Association of Module Completion and Acceptability With Participant Demographics and Distress Levels**

No association was detected between module acceptability and education, farm type, remoteness, age, internet use, or baseline psychological distress (Multimedia Appendix 1). Acceptability ratings were related to stress scores (\( \beta =0.14, 95\%\ CI 0.06-0.22; P=.001\)); the more stressful the events of the past month, the more satisfied participants were with the modules. Acceptability ratings were related to gender at \( P=.08 \), indicating a possible trend toward females finding the modules more satisfying than males.

No association was detected between the number of modules completed and gender, education level, farm type, remoteness, hours of internet use per week, or baseline psychological distress or stress exposure (Multimedia Appendix 2). There was an association between module completion and age, with older participants completing more modules (\( \beta =0.03, 95\%\ CI 0.00-0.06; P=.04 \)). Finally, there was no association between module completion and an individual’s average acceptability rating (\( r=-0.04; P=.52 \)) or their rating of the last module they completed (\( r=0.10; P=.28 \)).
Stage 8b: Usability and Acceptability Testing of Prototype (Qualitative Data)

Brief Phone Calls With Users Who Exited the Intervention Before Completion of All 5 Modules (to Find Out Why)

Table 6 summarizes the most frequently identified reasons for not completing a module (N=108). Most often, farmers said they were too busy or simply had not got to it yet (86/108, 79.6%). The next most common reason was that the content was not relevant to them (14/108, 13%) or that they had forgotten about it (8/108, 7.4%).

### Table 6. Reasons for not completing all 5 modules (N=108)².

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of mentions, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too busy or not got to it yet</td>
<td>86 (79.6)</td>
</tr>
<tr>
<td>Not relevant or helpful for me</td>
<td>14 (13)</td>
</tr>
<tr>
<td>I forgot or thought I had done it</td>
<td>8 (7.4)</td>
</tr>
<tr>
<td>Technical issues: user end</td>
<td>7 (6.5)</td>
</tr>
<tr>
<td>Module took a while or too long</td>
<td>7 (6.5)</td>
</tr>
<tr>
<td>Repetitive questions</td>
<td>7 (6.5)</td>
</tr>
<tr>
<td>My health</td>
<td>5 (4.6)</td>
</tr>
<tr>
<td>Technical issues: ifarmwell</td>
<td>4 (3.7)</td>
</tr>
<tr>
<td>Forgot password or reset issues</td>
<td>2 (1.9)</td>
</tr>
</tbody>
</table>

²Some participants gave ≥1 reason.

Detailed Phone Interviews With Purposively Sampled Group of Users

In total, 4 broad themes and 25 subthemes were identified and are outlined in Table 7. The themes included the following: using ifarmwell was a positive experience, value for themselves but unsure how best to recommend to others, areas for improvement, and context. The findings generally indicated that users found ifarmwell easy to use and navigate, relevant, credible, and necessary, particularly because of the tough drought conditions that many farmers were experiencing at the time of data collection. Farmers generally liked the structure of the modules and the time provided between modules to practice strategies. They also consistently reported that the language, videos, and cartoons were appropriate, the email or text reminders were helpful, and they valued the opportunity for self-reflection and the anonymity and privacy of the resource. Findings regarding areas for improvement included using even more farmer-focused language, improving the sound quality of the audio files, and including additional reminder SMS text messages to address forgetfulness. Module 3 was also identified as too long, and the values exercises it contained were found to be difficult for people who had never considered this type of value clarification exercise before. The need to double click to answer questions on iPads and iPhones was also something that users said they needed the website to remind them to do.
<table>
<thead>
<tr>
<th>Themes</th>
<th>Example quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Using ifarmwell was a positive experience</strong></td>
<td>“When I started it off I thought, ‘These guys have been reading my mind or watching me,’ because it seemed very pertinent, very pertinent. But also, just the fact that there’s no shame. I don’t have to be ashamed of the fact that I can’t help the things I can’t help. That’s a very empowering and liberating sort of a thing, so I got that from you.” [female, 56 years, VIC, sheep and/or cattle property]</td>
</tr>
<tr>
<td>Easy to use and navigate</td>
<td>“Very usable, I was really impressed with the usability of it, it was very simple and I am quite computer literate but I can imagine that someone that perhaps wasn’t so computer literate, the layout and the sequential nature of it, was pretty good.” [female, 61 years, VIC, sheep and/or cattle property]</td>
</tr>
<tr>
<td>Relatable and relevant to farmers</td>
<td>“Yes, if it was just for the ordinary person, which would be, of course, an urban person, it would be very, very different. I’m very grateful it was something focused on farmers because it just - well, it personalised it. It understands what’s going on.” [male, 64 years, WA, grain, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>“Because you’ve structured it for farmers. We’re very down-to-earth people, and I think some of these other courses weren’t down-to-earth enough. So, your language is being appropriate, your contents are appropriate, illustrations are brilliant.” [female, 66 years, NSW, horticulture, market garden, and/or fruit growing]</td>
</tr>
<tr>
<td></td>
<td>“But like it definitely - yeah read as something that was relevant from a rural perspective and approachable I guess, didn’t strike me as someone in an office in Sydney telling us how we should be dealing with the issues of rural mental health or whatever. Like it came across as real.” [female, 33 years, QLD, sheep and/or cattle property]</td>
</tr>
<tr>
<td>Content credible and well-developed</td>
<td>“I’m just trying to think of the—yeah, I think everything—well, there was nothing in there that I feel was irrelevant or inappropriate at any point.” [male, 42 years, NSW, grain, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>“Yes, I think it’s quite credible. Everything that was written there it was well written, it was easy to understand. I know it said if you need help call Lifeline. There was, that was on there, so yes, it was quite good I thought. Definitely, it looked good. You have obviously spent a lot of work on it. I found it good.” [female, 31 years, VIC, grain, sheep and/or cattle property]</td>
</tr>
<tr>
<td>Appropriate language and explanations</td>
<td>“It was good, it was simple. Not too simple that made you feel like, dumb or anything. They didn’t have big words either that you need to look up. So yes, it was quite appropriate I think for the demographic that you’re trying to target.” [female, 31 years, VIC, grain, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>“Just the way you chose the words, you didn’t make it more complex than it needed to be and you didn’t use technical jargon, it was very simple, everyday language.” [female, 61 years, VIC, sheep and/or cattle farm]</td>
</tr>
<tr>
<td>Videos were relatable, accessible, good quality, and useful</td>
<td>Again, I thought they (the videos) were really good because they are relatable and they are real. [female, 55, VIC, Dairy farm]</td>
</tr>
<tr>
<td>Appropriate use of photographs and cartoons</td>
<td>“Yeah, so happy again with those because they really, I think they were chosen well to reflect the environment of the people that you’re hoping to reach. You know, kept things within that framework, so yeah, no, absolutely happy with all of those.” [male, 25 years, WA, grain, sheep and/or cattle farm]</td>
</tr>
<tr>
<td>Modules were presented in a logical sequence</td>
<td>“I liked the way that it was broken up into different modules so that you were able to look at a section, do the skills and be exposed to some new skills and then have time to consolidate and think about that. For me, that’s a really good way to learn new skills, rather than just looking at something on that and then going ‘oh that was interesting’, it sort of was dribbled out a little bit over a period of time and I found that a really useful format for developing a structure for reflecting on how you deal with life and I think that’s a really useful way for a lot of farmers too.” [female, 61 years, VIC, sheep and/or cattle farm]</td>
</tr>
<tr>
<td>Valued time to implement strategies between modules</td>
<td>“It was good because it gave you a chance to practice or think about some of the things that you’d discovered, and then—without overloading you, and then you had another follow-up at the next step. I really liked the way that it did that. Like I said, it made it a much more sustainable sort of process.” [male, 36 years, VIC, viticulture]</td>
</tr>
</tbody>
</table>
### Themes

| Email, text, or voicemail, exercises, reminders were helpful | “Yeah, so as much as I hate enlisting in something and they keep bugging me all the time, I thought the texts as well as the things were good, particularly when you’ve got a fortnight between stuff. Yeah, I thought that was really good.” [male, 42 years, SA, grain, sheep and/or cattle farm] |
| Practical strategies | “It was good...It was quite practical in the way it was presented, the information was presented...Some of the examples that were presented and things like that were something you can easily identify with. It didn’t go into too much detail.” [male, 36 years, VIC, viticulture] |
| Using ifarmwell facilitated self-reflection | “I think—it took me a long time to identify and realize that I needed to do something with my mental health. It takes a lot to go forward and speak to someone, so being able to go through those modules on your own and identify where you need—you might need some help or even just identifying a few things that you can do for yourself, I think that probably suits farmers or anyone I’ve ever dealt with at work. I think being able to do something on your own to start with and get a [00:08:38], if this gives you a bit of information, really, to—then if you want to go to someone, you can say, ‘Look, this is what I think I need help with.’ That’s where I really struggled. I didn't know—I didn't really know what to—if I was going to go and talk to someone, I didn't really know what to say. But now I—having gone through those modules, it really highlighted for me.” [male, 36 years, VIC, viticulture] |
| A necessary and timely resource | “I hope it rolls out because to me it’d be a fair loss if it did not keep going—for sure. So I suppose that means that I better swallow my pride and actually tell someone about it.” [male, 42 years, SA, grain, sheep and/or cattle farm] |
| Appreciated the opportunity to add tools to Toolbox and refer back to summary sheets over time | “I think it’s a good program, you’d say, I suppose. It’s probably what we need right now too.” [female, 31 years, VIC, grain, sheep and/or cattle farm] |
| Able to understand strategies and apply them to life | “So no, I hope it doesn’t disappear because I think there is a definite need there.” [female, 33 years, QLD, sheep and/or cattle farm] |
| Appreciated privacy and anonymity | “I quite liked the way you could put stuff in your Toolbox. You could find those things that were potentially going to work for you and put them somewhere so you can refer to them later or coming back to them.” [male, 36 years, VIC, viticulture] |
| Willing to recommend to peers | “Yeah, I thought that was really good...It was quite practical in the way it was presented, the information was presented...Some of the examples that were presented and things like that were something you can easily identify with. It didn’t go into too much detail.” [male, 36 years, VIC, viticulture] |
| Value for themselves but unsure about how best to recommend to others | “That’s where I'm doing most of my promoting. I say to the girls—not just girls, to all the people, ‘This ifarmwell thing, it was a brilliant idea because this helps. It’s particularly tailored for farmers.’” [female, 56 years, VIC, sheep and/or cattle farm] |
### Themes

<table>
<thead>
<tr>
<th>Example quote</th>
<th>Areas for improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Yeah I would definitely and I actually thought, while I was going through, there is probably—well, I actually think it would do my partner a lot of good to do it as well, but I haven’t quite worked out how to encourage him to do that. But I definitely would given the right type of circumstances” [female, 33 years, QLD, sheep and/or cattle farm]</td>
<td>More farmer-focused language • “You didn’t put enough farmers’ language in there.” [male, 65 years, SA, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Improve the sound quality of audio files • “Actually one thing that was a bit of a problem was the, when there was meditations that, the girl that was doing the meditations, her voice was quite low and I couldn’t turn it up. So that was a bit of an issue. I could get through with it but it was, that was something that I did notice” [female, 62 years, TAS, horticulture, market garden, and/or fruit growing]</td>
</tr>
<tr>
<td></td>
<td>Include additional reminders • “Maybe more reminders. I know for me I obviously signed up and I suppose people that do sign up to do these things do have the intent to do it. Like everything, you sort of get emails from here, there and that’s just life these days and that’s just the way it is. But I would appreciate obviously another reminder being like ‘Come on!’” [female, 23 years, NSW, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Shorten module 3 • “I think that one [module 3] took me the longest time, actually. I did—I think a lot of those things were relevant, and then after a while I dragged and dropped all these things and I began to regret it a little bit, because it took so much time to sort it out and comment on each one. I think that’s what happened, so it was a bit lengthy.” [male, 64 years, WA, grain, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Remind users to double tap to select answers on iPads and iPhones • “The only thing—like there was a note about it was that you had to double tap because I did a fair bit of it on my phone and...a couple of times like you would do your multiple choice and I would have to go back because it would say you haven’t answered it. I’m like, ‘Oh, I did answer it.’ But just so obviously hadn’t but there was a note in there telling you what you had to do and that was fine but I would say that was more operator error than internet thing.” [female, 33 years, QLD, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Context Farmers are time poor • “I am thinking about—from it personally but I am also thinking about it in terms of professionally and how I would perhaps recommend something like that to farmers that I am working with as well and I think that the fact that it’s not a very time consuming thing, each module means that you can just do a little bit at a time and you can jump in and out of it, depending on what time requirements you have so the overall structure I thought was terrific from that perspective.” [female, 61 years, VIC, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Mental health stigma • “I guess probably a lot of farmers probably baulk when they hear something about mental health, feelings and emotions and that sort of thing” [male, 36 years, VIC, viticulture] • “I think it’s a really good idea because it’s—farmers are very proud people. They won’t always go and seek help. But this is kind of non-threatening. They don’t have to talk to anybody if they don’t want to.” [female, 57 years, TAS, grain, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Drought • “We can’t do anything about the weather. We can’t change it. I haven’t got any feed.” [female, 71 years, NSW, sheep and/or cattle farm] • “And the other things I liked about it was just that you are farmer-orientated, which is totally different to any of the other help—beyondblue, Black Dog, they’re all just for general people but farming situations are particularly unique and your ‘ifarmwell’ tapped in to that—so the idea that drought or cattle prices that you can’t influence and, more importantly, succession.” [female, 56 years, VIC, sheep and/or cattle farm]</td>
</tr>
<tr>
<td></td>
<td>Women are perceived as most likely to use and recommend • “I think, the wives, I reckon the wives would be more likely to be interested in it.” [female, 62 years, TAS, horticulture, market garden, and/or fruit growing]</td>
</tr>
</tbody>
</table>

---

**VIC**: Victoria, Australia.  
**WA**: Western Australia, Australia.  
**NSW**: New South Wales, Australia.  
**QLD**: Queensland, Australia.  
**SA**: South Australia, Australia.  
**TAS**: Tasmania, Australia.
Stage 9: Iterative Design Improvements

In response to the findings detailed above, several changes were made to the website. To improve clarity and brevity, minor wording changes and reductions in the text were made in all modules. Audio recordings were professionally rerecorded to improve quality. Edits to the text were also made to acknowledge that accessing a GP can be difficult for those in rural areas, that module 1, in particular, was very long because of the pre-evaluation questionnaires (but that subsequent modules would involve less reading and more activities), and that questionnaires were standardized and only included for the purposes of website evaluation (not part of the intervention itself; eg, cognitive fusion). To improve usability, additional reminders were included about the inability to go back and the need to double tap responses if using an iPad or iPhone. A Things to remember when using this website page was added to emphasize these key messages. The save and continue button was also made more prominent. To improve relevance, additional images and rotating banners were included on the home page to reflect the broader range of demographics of users accessing the website. To improve adherence, additional SMS text messaging reminders were added 7 days after registration if module 1 was not completed and 28 days after the preceding module was completed if the next module was not started for modules 2 to 5 (Figure 2). Module 3 was shortened by removing 1 value clarification exercise that gave users feedback on values they may not be living consistently with (based on their answers to a questionnaire; Table 3). The revised module 3 retains an exercise asking users to select values that are very important to them, think about whether these values drive their behavior and decision-making, and how they might plan to live more consistently with these values in the future.

Discussion

Principal Findings

This paper outlines the process of integrating evidence from the literature and consumer and expert advice to create a resource that is informed by evidence and perceived as acceptable and relevant by its users. A strength of this intervention development process was the clear, iterative methodology that allowed the integration of different types of knowledge at each step. This involved the synthesis of evidence from prior research and intervention mapping to identify key determinants of behavior change, relevant behavior change and engagement strategies, and the involvement of farmers as co-designers throughout the process to ensure the acceptability of evidence-based strategies. In particular, the farmers’ feedback was used to inform the initial design of the website, amend the prototype before launch, inform the timing of the launch, and update the intervention following acceptability and usability testing of the prototype. At all stages, farmers’ feedback was prioritized and integrated with research evidence and expert opinions. These approaches enabled us to develop a resource that reflects the unique farming culture, is built on evidence-based approaches to mental health and well-being, demonstrates an understanding of the audience for which it was intended, and as detailed in this paper, was found to be acceptable.

More specifically, the acceptability and usability testing of the prototype that included both quantitative and qualitative components and farmers from a variety of Australian states and farm types, found that once people started a module, most completed that module. Approximately 83.6% (117/140) of users starting module 1 completed module 1, and all people who started modules 4 and 5 completed them. Importantly, acceptability with the previous module was not found to predict whether a user went on to complete the next module, which aligns with the qualitative feedback from people who did not complete every module that their main reason for not progressing was Too busy/not got to it yet. Overall, 37.1% (52/140) of people who started module 1 completed the entire 5 module intervention. Comparatively, recent studies have shown a wide variation in the rate of adherence and attrition to web-based interventions for mental health between 2% and 83% [72]. Other studies have reported that approximately 75% do not use mHealth apps more than once after installation [73]. Pleasingly, the present intervention was found to be most acceptable to those who needed it most (ie, those who were most highly distressed when they started module 1) rather than those who were most educated. These high levels of acceptability are significant, given the aforementioned reluctance of farmers to seek help [14], engage with resources targeting their mental health [37], and their general perception that existing services are not designed for them [24]. The intervention also aims to help farmers identify when and how to seek professional help and highlights the role of their local GP. In turn, this may prevent the development of severe mental health problems and facilitate access to treatment at an earlier stage, thereby minimizing the intensity of interventions required and reducing both social and treatment costs. Findings from the qualitative interviews with noncompleters (N=108) to find their reasoning for ceasing participation, also met calls for more research to aid the understanding of engagement in web-based interventions [74] and may be used to inform the inclusion of strategies for improvement in future interventions.

The only comparable farmer-focused well-being website is the aforementioned Scottish CCBT Living Life to the Full, which includes personalized support emails in addition to computerized modules [23]. That trial found that of those who logged on (N=35), only 5 (14%) completed the 5 core modules, which is much less than the 37.1% (52/140) reported in this study. Bowyer et al [23] noted that rates of attrition in their study with farmers (73.2%) were much higher than those experienced when they tested a very similar intervention with other population groups (26%-27% attrition) [75,76], reinforcing the notion that the farming population is particularly difficult to engage in health and well-being–focused interventions.

Although acceptability with the ifarmwell modules was generally high, along with the interview comments, they did highlight some areas for improvement. Following the acceptability testing reported in this paper, the website was adapted to address any concerns and improve the user experience. Changes included shortening a module, improving the quality of audio recordings, and incorporating additional SMS text message reminders, which demonstrates the value of adaptive design in building a resource that is responsive to user needs to double tap responses if using an iPad or iPhone.
experiences. This aligns with the person-based approach by continuing to incorporate user feedback after live testing of the intervention [34], which is a strength of this work as it allows interventions to be responsive to the needs of the audience while remaining publicly available. The need to ensure that modules are as short as possible (or can be easily stopped and recommenced) is important for other farming-focused intervention developers to keep in mind. Our finding that farmers lack the time to engage in web-based interventions aligns with findings that more than half of the Australian farming population work ≥50 hours per week, compared with just 16% of the rest of the working population [77].

Limitations
The sample was limited to those farmers who self-selected to take part in the website evaluation and may not be representative of the wider farming community [78]. Another limitation of this research is that it was not clear to many users that the questionnaires used for evaluation purposes were not part of the intervention itself, which may have contributed to the perception of module length and negatively affected user acceptability. A yellow background was used behind the evaluation components; however, in the future, this delineation should be made even clearer, possibly by having users access the questionnaires via a separate window.

Further Research
We have demonstrated that a co-designed website is usable and acceptable to farmers, and many of the lessons from this research may be applied to the development of future farmer-focused interventions. However, further research is needed to systematically test the effectiveness of this intervention and examine the psychological mechanisms that facilitate changes (or otherwise) in outcomes. In the case of ifarmwell, analyses should specifically examine whether key ACT processes (Table 3) are influenced by the intervention and, if so, how they relate to any changes in distress and well-being outcomes. This would not only inform further refinements to the ifarmwell website but also help progress important gaps in knowledge about psychological mechanisms in the field of ACT [43,51].

Conclusions
This paper describes the first web-based intervention co-designed with farmers to help them adopt coping strategies to better manage their stress by accepting things beyond their control and living according to their values, regardless of the circumstances they face. Importantly, this paper outlines the value of a co-design approach in facilitating the development of interventions that are centered on evidence-based therapeutic approaches, that also appeal to audiences who are typically reluctant to seek help for mental health problems. It also details a comprehensive, successful website development and acceptability testing process, which may inform the development of future web-based interventions for difficult-to-reach populations.

Acknowledgments
The authors gratefully acknowledge the farmers who took part in this research and assisted with the development and testing of the intervention, and Mrs Margaret Mc Gee for her assistance with supporting this project.

This project was funded by the NAB Foundation, University of South Australia, Freemasons Foundation Centre for Men’s Health, and the Freemasons Foundation.

Authors’ Contributions
KMG was involved in the conceptualization, methodology, funding acquisition, supervision, project administration, data collection, data analysis (qualitative), and writing of the original draft. GS was involved in the data analysis (quantitative) and writing of the original draft. JD was involved in the conceptualization, methodology, funding acquisition, and writing—review and editing. ADV was involved in the conceptualization, methodology, data analysis (quantitative), and writing—review and editing. CES was involved in the conceptualization, methodology, and writing—review and editing. SB was involved in the conceptualization, methodology, and writing—review and editing. AB was involved in the project administration, data collection, data analysis (qualitative), and writing—review and editing. NH was involved in the project administration, methodology, data collection, and writing—review and editing. DT was involved in the conceptualization, methodology, and writing—review and editing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Linear mixed model assessing the relationship between module number, demographic and distress variables on module acceptability rating.

[DOCX File, 15 KB - humanfactors_v9i1e27631_app1.docx]

Multimedia Appendix 2
Univariable and multivariable linear regression models predicting highest number of modules completed.
References


38. Gunn et al, JMIR HUMAN FACTORS 2022 | vol. 9 | iss. 1 | e27631 | p.73 https://humanfactors.jmir.org/2022/1/e27631


71. Perez S. Nearly 1 in 4 people abandon mobile apps after only one use. TechCrunch. 2016. URL: https://techcrunch.com/2016/05/31/nearly-1-in-4-people-abandon-mobile-apps-after-only-one-use/ [accessed 2020-11-25]


Abbreviations

- ACT: acceptance and commitment therapy
- CCBT: computerized cognitive behaviour therapy
- GP: general practitioner
Educators' Perspectives on Integrating Technology Into Sexual Health Education: Implementation Study

Martha J Decker¹, MHS, DrPH; Salish Harrison², BA; Melisa Price³, MPH; Abigail Gutmann-Gonzalez², MPH; Jennifer Yarger³, PhD; Rachel Tenney⁴, MD

¹Department of Epidemiology and Biostatistics, Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, San Francisco, CA, United States
²Philip R. Lee Institute for Health Policy Studies, Bixby Center for Global Reproductive Health, University of California, San Francisco, San Francisco, CA, United States
³Clinical and Translational Science Center, University of California, Davis, Davis, CA, United States
⁴School of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Martha J Decker, MHS, DrPH
Department of Epidemiology and Biostatistics
Philip R. Lee Institute for Health Policy Studies
University of California, San Francisco
550 16th Street, Second Floor
San Francisco, CA, 94158
United States
Phone: 1 (415) 476 3095
Email: Mara.Decker@ucsf.edu

Abstract

Background: In the last decade, the use of technology-based sexual health education has increased. Multiple studies have shown the feasibility of technology-based interventions, while a subset has also shown efficacy in improving youths’ sexual health outcomes such as increased condom use and knowledge. However, little is known about health educators’ experiences in integrating technology to augment sexual health curricula.

Objective: The purpose of this study was to assess the perceptions and experiences of health educators regarding the incorporation of technology into a sexual health education program designed for underserved youth in Fresno County, California, and to identify facilitators and challenges to incorporating technology into the in-person curriculum.

Methods: This implementation study used data collected as part of a cluster randomized controlled trial to evaluate In the Know (ITK), an in-person sexual health education curriculum that includes technology-based content, such as a resource locator, videos, and games, which can be accessed through a mobile app or website. Data from implementation logs from each cohort (n=51) and annual interviews (n=8) with health educators were analyzed to assess the health educators’ experiences using the technology and adaptations made during the implementation.

Results: The health educators reported that technological issues affected implementation to some degree: 87% of the time in the first year, which decreased to 47% in the third year as health educators’ familiarity with the app increased and functionality improved. Technology issues were also more common in non–school settings. Successes and challenges in 3 domains emerged: managing technology, usability of the ITK app, and youth engagement. The health educators generally had positive comments about the app and youth engagement with the technology-based content and activities; however, they also noted certain barriers to adolescents’ use of the mobile app including limited data storage and battery life on mobile phones.

Conclusions: Health educators require training and support to optimize technology as a resource for engaging with youth and providing sensitive information. Although technology is often presented as a solution to reach underserved populations, educational programs should consider the technological needs and limitations of the participants, educators, and settings.

International Registered Report Identifier (IRRID): RR2-10.2196/18060

(JMIR Hum Factors 2022;9(1):e31381) doi:10.2196/31381
KEYWORDS
adolescent; sex education; technology; mobile app; implementation; California; health educator

Introduction
The use of technology-based sexual health education programs aimed at reducing sexually transmitted infections and unplanned adolescent pregnancy has increased over the last decade. Teaching with technology can be defined as any type of educational process that incorporates digital technology tools such as television, computers, tablets, smartphones, mobile apps, online educational games, or online collaborative learning environments to advance student learning [1].

The use of digital technologies in sexual health education programs has increased for multiple reasons. Some data suggest that youth access to the internet and web-based content has become nearly ubiquitous. A Pew Research report showed that 95% of adolescents aged 13-17 years had access to a smartphone in 2018 with almost 45% reporting being online on a “near-constant” basis and 90% going online multiple times per day [2]. Using technology for entertainment and information seeking may be particularly appealing in adolescence, and technology may also help to reinforce adolescent developmental growth through exploration and social connection [3]. Digital technology may also help alleviate student and teacher embarrassment, which is common when discussing sensitive subjects during sexual health classes [4]. In addition, technological tools may be able to reach and educate marginalized youth who lack access to quality and inclusive sexual health education in their schools [5,6]. However, recent research shows that a digital divide persists even among young people [7]. For example, in 2019, low-income adolescents were less likely to own laptops and smartphones than high-income adolescents (36% vs 54% and 74% vs 89%, respectively) [8].

Prior research has demonstrated that youth have favorable opinions of technology-based sexual and reproductive health interventions [9-12]. Some studies also have shown that interventions that incorporate technology were effective in improving youths’ sexual health outcomes, such as condom use, abstinence, sexual health knowledge, and safer sex norms [13-15]. However, a previous review of sexual health education apps found that most lacked comprehensive sexual health content and had limited interactivity, highlighting the unmet potential for this type of platform [16].

Despite this increase in digital sexual health interventions, little is known about health educators’ experiences delivering sexual health interventions that incorporate technology-based components. Previous research on technology in general educational programming found that health educators’ lack of confidence and perceived value of the technology can be barriers to integration [17,18]. One implementation evaluation of an online sexual health program in the Netherlands reported that while teachers appreciated the interactive content, they often needed to adapt the materials based on classroom dynamics, and some found transitioning between web-based and classroom teaching challenging [9]. Coaches in a sports-based HIV program in South Africa, which included text messages as part of the intervention, identified students’ shortage of cellular data as the primary challenge [19]. With the growing interest in online and technological approaches to education, it is critical to learn from the experiences of health educators in incorporating technology to ensure that digital content is a viable resource for engaging with youth and improving sexual health knowledge and behavioral outcomes.

The purpose of this study was to assess the perceptions and experiences of health educators regarding the integration of technology into a sexual health education program called In the Know (ITK) and to identify facilitators and challenges to incorporating technology into the in-person curriculum. These results can help future program developers and health educators anticipate and mitigate common issues with technological integration and promote best practices.

Methods
Intervention Overview
ITK was developed by and for adolescents aged 13-19 with a goal of increasing use of contraceptive and clinical health services [20]. Adolescents representing diverse priority populations engaged in a user-centered design process to help create the program’s content and digital components [21,22]. The curriculum is based on a positive youth development approach, which promotes personal strengths and healthy development through supportive opportunities and experiences [23,24].

The program was developed to be inclusive and to address the needs of homeless and unstably housed youth; youth of color; and lesbian, gay, bisexual, transgender, and queer or questioning (LGBTQ+) youth. ITK combines 6 hours of in-person sexual health education with technology-based content to provide the skills, information, and resources necessary to improve the sexual and reproductive health and overall well-being of adolescents.

The intervention is divided into three modules: (1) sexual health and contraceptive use; (2) healthy relationships; and (3) educational and career success. Health educators incorporated different technology-based components in each module, such as videos, online goal setting, career opportunities, and geo-location of local services. Some content was “gamified” using Kahoot, a game-based learning platform, and app-based quizzes and activities to earn points. Health educators concluded each module with a guided activity on the app and then assigned a task for the youth to complete outside of class. Youth could also sign up to receive text message reminders of key content and personal goals. These tools as well as additional resources and quizzes were available through a downloadable app or website, enabling youth to access the information outside of the in-person sessions. Health educators provided tablets with the app previously installed for use during the in-person sessions, though the participants also were encouraged to download the app on their mobile phones. The health educators helped to
The health educators received training on the curriculum, classroom management, and the technological components prior to implementation. This included in-person trainings and “teach backs” as well as shorter refresher trainings throughout the implementation period. The health educators reported technical issues about the app to the website developers. In addition, the developers updated features and replaced broken links over the course of the program. The researchers, health educators, and the app developers met biweekly to discuss any implementation challenges and adaptations.

**Setting and Participants**

ITK was implemented in 51 cohorts (groups) with 559 youth at 36 youth-serving agencies representing a variety of settings where youth receive services or activities in Fresno County, California. The health educators traveled to the sites of the participating agencies for implementation, which included school and after-school settings, employment and training sites, youth development centers, clubs, foster care sites, housing authorities, tribal agencies, and LGBTQ+ programs. The majority of participating youth were Latino (70% [n=381]) with a mean age of 15.5 years (SD=2.07). Almost all of the participants owned or shared a smartphone (89% [n=480]), and 86% (n=469) had access to the internet in their homes.

Over the 3 years of implementation, a total of 6 health educators implemented ITK, with an average of 3 health educators per year. The health educators had a range of educational backgrounds, prior teaching, or training experience, and were comfortable with technology. This varied from 1 educator with over 6 years teaching comprehensive sexual health education to 2 educators with no prior experience in sexual or reproductive health and limited familiarity with technology; 2 other health educators had at least 2 years of experience implementing sexual health education in similar settings. Moreover, 2 health educators were male, and all lived in Fresno County.

**Data Collection**

This implementation study used data collected as part of the cluster randomized controlled trial [20]. Due to the complexity of the intervention being evaluated, a better understanding of the contextual factors, including the technology and in-person implementation, can help to improve future interventions and interpret the intervention’s outcomes [25]. Process data from implementation logs and annual interviews with health educators were collected to assess fidelity to the intervention and to promote ongoing quality improvement.

**Implementation Logs**

Health educators completed an implementation log after delivering the program to each cohort. A cohort is a distinct group of participants receiving ITK at a specific time, such as a classroom of students. Each log consisted of 6 main sections: physical space, teaching methods, learning environment, youth participation, classroom management, and technology. The health educators were encouraged to comment on any contextual factors or circumstances that facilitated or hindered program delivery for specific activities or for the entire cohort. Each log also included a closed-ended question, “Thinking about what happened across all of the modules of this cohort, how often did technology issues impact implementation?” with the response options being all, most, some, or none of the time. At the end of each cohort, the health educator uploaded the completed log to Box, a secure online file management system. The researchers reviewed the implementation logs for completeness and accuracy after submission and debriefed with the implementing health educator.

**Health Educator Interviews**

The researchers conducted annual interviews near the end of each school year with the health educators for 3 years. Due to staffing changes over that time, 2-3 health educators were interviewed each year, with 2 of the health educators interviewed twice. Topic areas included implementation experiences, youth reactions, perspectives on the digital technology components, and recommendations. The interviews were conducted in a private office and averaged 53 minutes in length. All interviews were audio recorded and transcribed verbatim. Health educators received a US $20 gift card in appreciation of their time.

**Analysis**

This study used a modified form of grounded theory in which a set of potential concepts were identified and coded, and additional themes were inductively identified from the data [26]. The qualitative analysis was guided by structural themes based on key areas of research interest, such as technology use, emerging themes from the review of transcripts, and the open-ended responses in the implementation logs [27]. This mixed coding system combined an initial list of codes using the main research questions and additional codes that were added based on further review [28].

One researcher coded all transcripts, while another double-coded a subset and reviewed coding for intercoder consistency. The coded interviews had an interreliability score of 0.80. The research team met regularly to review the coding process, clarify codes, and update the codebook. As needed, the researchers reviewed the quotes that were coded differently and jointly agreed to their coding. The codes were analyzed for patterns, with relevant themes extracted. The findings were also compared by year and by health educator to assess if experiences varied over time or by person. The qualitative coding was conducted using Dedoose, version 8.0.35 (SocioCultural Research Consultants, LLC) [29].

The responses to the closed-ended question regarding the frequency of technology-related interruptions were extracted and summarized using Stata 16 (Stata Corp). We used the Fisher exact test to compare the responses by whether the cohort received the program in the first year of implementation and in a school setting. One-sided $P$ values are reported.

**Results**

**Technology Issues During Implementation**

Implementation logs were completed for all 51 in-person sessions of ITK conducted between October 2017 and February
During the first year of implementation, the health educators reported that technology issues affected implementation to some degree of the time in 7 out of 8 cohorts (87%) with that amount decreasing over the next 2 years, to 11 cohorts out of 19 (58%) and 7 cohorts out of 15 (47%), respectively (Figure 1; note that there were missing responses from 9 implementation logs in year 1 since one question on technology issues was added later). When calculated with the Fisher exact test, the difference between the first year and subsequent years was only marginally significant ($P=0.08$) due to the small sample size. The cohorts implemented in non–school settings such as in group homes or community-based organizations were much more likely to have technology issues than those in school settings; 14 out of the 19 (74%) non–school setting cohorts experienced technology issues compared to 11 out of 23 (48%) of cohorts implemented in a school setting ($P=0.05$).

Figure 1. Percentage of time when implementation was affected by technology, by year (n=42).

Successes and challenges emerged in 3 key domains: managing technology, usability of the ITK app, and youth engagement. Managing technology included issues related to meeting the technological requirements and administrative needs for implementation during the in-person ITK sessions, such as device compatibility, internet access, and availability of necessary technology hardware. The topics related to the ITK app’s usability were those specific to the content and functionality of the app, the integration of the app into the in-person ITK curriculum, and the participants’ use of the app. Youth engagement referred to how integrating technology into the curriculum affected the participants’ focus and engagement during in-person implementation. Note that many of the issues overlapped; for example, challenges with internet connectivity limited access to the app’s content, which then affected youth engagement. Table 1 summarizes the successes and challenges within these 3 domains.
Table 1. Key successes and challenges of integrating technology into in-person sexual health education, by domain.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Successes</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing technology</td>
<td>• Implementation sites with audiovisual projection devices present (eg, TV, projector, speakers)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Bringing mobile Wi-Fi hotspot</td>
<td>• Significant preparation time required</td>
</tr>
<tr>
<td></td>
<td>• Providing tablets for classroom use</td>
<td>• Packing tablets, Wi-Fi packs</td>
</tr>
<tr>
<td></td>
<td>• Implementation site lacked necessary hardware</td>
<td>• Ensuring all devices were charged and functional</td>
</tr>
<tr>
<td></td>
<td>• Technology issues caused delays or omission of instruction</td>
<td>• Implementation site lacked necessary hardware</td>
</tr>
<tr>
<td></td>
<td>• Tablets freezing and needing to restart</td>
<td>• Technology issues caused delays or omission of instruction</td>
</tr>
<tr>
<td></td>
<td>• Internet connectivity issues</td>
<td>• Tablets freezing and needing to restart</td>
</tr>
<tr>
<td></td>
<td>• Inability to connect to the Internet led to alternative instructional methods</td>
<td>• Internet connectivity issues</td>
</tr>
<tr>
<td>ITK (In the Know) app</td>
<td>• Positive response to online content and resources</td>
<td>• Use of hard copies instead of digital content</td>
</tr>
<tr>
<td>usability</td>
<td>• App used as reference for local services</td>
<td></td>
</tr>
<tr>
<td>Youth engagement</td>
<td>• Use of tablets increased youth engagement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Certain digital content resonated well with youth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Youth playing on electronic devices led to distractions</td>
<td></td>
</tr>
</tbody>
</table>

Managing Technology

Implementation of ITK involved managing multiple technological devices and administrative requirements such as connecting program tablets and participant smartphones to the internet and projecting digital content on a screen. The health educators noted that implementation was easier in sites that had the necessary audiovisual projecting devices, such as a monitor and projector and internet access. The health educators consistently reported the challenges associated with using technology during implementation, though this decreased in frequency each year. An inability to access high-speed wireless internet was the most commonly described technology issue reported. The health educators adapted to this issue by bringing their own mobile Wi-Fi hotspots to the sites. Other common technology issues included tablets freezing or crashing during use, the lack of audiovisual projecting devices, broken web links to external online content, and youth forgetting log-in information. One health educator described common experiences with the technology:

> As much as you rely on it and as great as it is, sometimes the links aren’t working, the buttons aren’t working, the screen goes blank, and you’re pressing the button and nothing’s working. Then you have to restart it. [Interview, Year 1]

These technology-related issues caused delays and required health educators to adapt how they delivered the program, both ad hoc and while preparing for future implementation sessions. One health educator described an example of an ad hoc adaptation as such:

> I had to use downloaded version of materials due to internet connections. Students were really excited for Kahoot [online learning platform] but, unfortunately, the game was not showing the possible answer to the students, and they could not participate the way it is usually played. I ended up reading the questions out loud and had the youth raise hands when the answer sounded correct. [Log, Year 2]

The health educators provided tablets for participants to access the app if they did not download it on their personal phones. While this increased access to the materials, managing the tablets required significant planning and preparation time, as health educators needed to ensure that all electronic devices were charged and functioning properly. One health educator gave the following explanation:

> We always get our materials ready... very important is coming to make sure all the tablets are charged... so we don’t have any delays the next day. [Interview, Year 3]

ITK App Usability

The health educators generally had positive comments about the content of the ITK app, particularly the interactive map linking youth to resources within their community. The health educators used these online resources with the participants to identify community clinics, counseling services, and help lines for youth and families experiencing violence. One health educator described the benefits of having resources consolidated on the app:

> I really do love the resources of [the app]. I always let the youth know, like, “Hey, in the app that we
talked about, you know, if you have any other questions, it'd be really great for you to go on the app and you can find basically anything... There's numbers, there's addresses..." Because a lot of, some of, them do have questions that sometimes I don't know how to answer right off the bat. So I say, “Hey look at the app,” so that's really great. [Interview, Year 3]

While youth could access the ITK website on a tablet during class, ITK originally was designed as an app for youth to download on their phones for later access. However, youth often faced challenges in using the mobile app including limited storage or data, limited battery life, lack of a personal cell phone number, difficulty remembering the required password, and sharing the phone with other family members. One health educator noted the following experience:

You have another group of kids that have phones, but there's always a reason why they don't want to download the app. They don't want to, they don't have space on their phone, their phone is like some crazy off-brand they can't find it. They don't have battery, phone is totally cracked, something about service, something about problems downloading the app. I don't know, it's different every time. [Interview, Year 2]

Youth also expressed reluctance to download the app due to confusion about the purpose and utility of the app as well as its connection to the in-person curriculum, resulting in a limited use of the app outside of the in-person sessions. Additionally, some youth did not have access to a smartphone at all, which not only prevented them from accessing the app outside of the in-person sessions, but also contributed to the participants feeling left out of the program. A health educator described the experience of 1 youth who was homeless as such:

One participant mentioned that she felt like she was being discriminated against because she didn't have a cell phone.... [Interview, Year 1]

Health educators also noted a lack of integration between the in-person elements of the curriculum and the ITK app. Because many of the ITK app features and activities were explained at in-person elements of the curriculum and the ITK app. Because of this, facilitators had to tell the youth from time to time to stop being on the tablets or they didn't participate in the program, don't get to experience the app or get to have the information on the app. But I think it would still really help if we can actually use the app for facilitating, and that the youth can go back on the app and look through things that we've talked about already, or stuff like that. [Interview, Year 3]

Youth Engagement

Overall, the health educators reported that youth were engaged and interested in the curriculum. They stated that participation and engagement increased among the youth when playing games with Kahoot, an online learning platform that allowed educators to gamify content delivery. One health educator explained it as such:

Oh, Kahoot. When it's working, it works great. Like when it's working, it's probably like the one thing that the youth get excited about, maybe because they already know what it is and they get to play it at school already. So they think right away, like, “Oh, yes, it's a game!” [Interview, Year 3]

Youth also responded particularly well to activities utilizing the O*Net OnLine website, a free online career exploration tool. However, health educators also noted that youth preferred participatory activities in general, whether technology-based activities and games or in-person activities such as role plays compared to lecture-based activities. One health educator described their experience as follows:

Sometimes we're not using the tablets or we're not doing like any kind of more of a group discussion. Like when there's listening in or something, or when I'm just asking them questions, it's really hard to, it's like school. It's like, okay, raise your hand or something like that. That's where I start to lose them. [Interview, Year 3]

Despite fostering interest and engagement, in some instances, the presence of electronic devices was distracting for some youth. One health educator described a common experience in an implementation log as follows:

Some youths had earphones plugged in the tablets, played games, or even took photos of themselves during the class time. Facilitators would walk around the room to ask the youth to stop playing with the tablets while a facilitator was presenting. Although facilitators had to tell the youth from time to time to stop being on the tablets, facilitators did their best to move the class along with fewer distractions. [Log, Year 1]

Another common youth engagement issue was the need to contextualize or personalize content for the participants. On almost every implementation log, the health educators noted instances where they had to reframe content or add explanations. For example, 1 health educator noted their role in providing supplemental information regarding a video on the biology of conception and pregnancy:

Youth did not seem to understand the video as far as the feedback that we got after when trying to discuss. Facilitator replayed the video and broke it down into different wording with each section. [Log, Year 3]
Discussion

Principal Findings

These findings illustrate some of the successes and challenges of integrating digital technology into an in-person sexual health education program from the critical perspective of health educators. As previous studies found, health educators commonly reported that technological issues such as connectivity and device compatibility affected implementation, which were not unique to sexual health education [9,17,18]. However, technological issues became less frequent over time, likely for 3 reasons. First, health educators gained experience and confidence in addressing common technological challenges, including making innovative adaptations or finding alternatives when technology malfunctioned. Second, additional training may have led to greater familiarity and comfort with the myriad of platforms and implementation strategies. Third, health educators provided ongoing feedback to the developers, which resulted in changes to certain technology features and problem resolution. The decline in technological issues demonstrates the importance of ongoing and iterative quality improvement processes and the need for sustained engagement by the app development team in any technology-based health education intervention. It also illustrates the need to ensure that health educators are comfortable and confident in using technology, either through prior experience or through training.

Despite the implementation challenges, the health educators held positive views about the value that technology added to the in-person education, particularly in engaging youth with the material. Overall, youth tended to be more involved when they actively interacted with the content, whether through the technology-based components or in-person activities. Technology may be one of many tools that can increase the interactivity of curricular content [12]. A review of a variety of computer-based technologies found that digital games had the most evidence supporting their use to increase student engagement [30]. Game-based activities were successful, supporting the evidence that well-designed gamification can increase student engagement and motivation, and demonstrating the potential for gamification of educational content [31,32]. While the digital content was generally well received by youth, health educators also noted that the technology-based activities were not fully integrated into the curriculum. This was similar to the findings by another study of an online sexual health education program from the critical perspective of health educators [12].

Although adolescents have widely adopted technology, our findings are reflective of research showing ongoing disparities in technology access and use at the individual, community, and institutional level [7,33]. While ITK was designed for youth in underserved settings including foster care and shelters, health educators were more likely to encounter technology issues such as lack of Wi-Fi and other hardware in non-school settings. This made the implementation of the technology components of the program more challenging [34]. Additionally, while most participants had phones, some had limited storage or shared the phone with other family members, making them less inclined to download or keep an app, particularly one that stored sensitive information. By contrast, other studies have found that youth appreciate the anonymity available through technology-based sexual health interventions [34].

While technology can enhance youth engagement and comprehension, this study highlighted the critical role of health educators who secure the hardware necessary for implementation, adapt the curriculum when technology fails, and contextualize and personalize digital content to meet the unique needs of the youth they serve. Other studies have demonstrated the importance of staff training, confidence, and self-efficacy for the success of efforts to integrate mobile technology into education [35,36]. Beyond technological competence, health educators also need the core capabilities in knowledge and skills to deliver effective, inclusive, and appropriate sexual health education, particularly when discussing sensitive sexual and reproductive health topics [37].

Limitations

A few limitations should be noted. The implementation log data is self-reported, so health educators may have underreported issues or interpreted a situation differently. However, these results also were consistent with annual interviews with the health educators. This study did not assess the prior experience or comfort level of the health educators with technology. In addition, because the ITK app changed over time in response to feedback and updates, some of the technical components or issues may have been resolved over time or varied by time period.

Conclusion

As more sexual health educational programs incorporate technology, they should consider the specific role and use of technological components from both a pedagogical and logistical standpoint. Developers should engage with youth and health educators when designing health curricula and apps to ensure that the content is integrated and promotes youth learning and engagement. App developers need to invest in usability testing and a system for reporting issues throughout implementation and iteratively update the product based on that feedback. Similarly, developers and organizations need to ensure that health educators have the training, confidence, and support necessary for successful implementation, including the curricular content, classroom management skills, and necessary technology.

Although technology is often presented as a solution to reach underserved populations, that premise is not yet fully realized. Educational programs considering the adoption or integration of technology should assess the potential needs and technological capacity of the participants and settings.
Acknowledgments

The authors thank our partners at the Youth+Tech+Health initiative of ETR, the staff at Fresno Economic Opportunities Commission, and all the study sites, as well as youth who participated in this study. This study was funded by the Department of Health and Human Services, Family and Youth Services Bureau, Personal Responsibility Education Program Innovative Strategies, grant 90AP2688-01-00. The funder had no role in the data collection, analysis, interpretation, or publication of this manuscript or future results.

Authors’ Contributions

MJD is the principal investigator of the study and is one of the primary authors of the manuscript. RT, SH, and MP coded and analyzed implementation logs and interviews. RT wrote the first draft of the introduction and results and edited the manuscript. MP provided training and oversight for fidelity monitoring and conducted health educator interviews. AGG provided support to this study’s evaluation activities, including health educator interviews. JY provided support to this study’s evaluation design. All authors reviewed and edited the manuscript.

Conflicts of Interest

None declared.

References


32. Teadt S, Burns JC, Montgomery TM, Darbes L. African American adolescents and young adults, new media, and sexual health: scoping review. JMIR mHealth uHealth 2020 Oct 05;8(10):e19459 [FREE Full text] [doi: 10.2196/19459] [Medline: 33016890]


Abbreviations

ITK: In the Know

https://humanfactors.jmir.org/2022/1/e31381 JMIR Hum Factors 2022 | vol. 9 | iss. 1 | e31381 | p.84 (page number not for citation purposes)
Using Health Concept Surveying to Elicit Usable Evidence: Case Studies of a Novel Evaluation Methodology

Alex Mariakakis, PhD; Ravi Karkar, PhD; Shwetak N Patel, PhD; Julie A Kientz, PhD; James Fogarty, PhD; Sean A Munson, PhD

1Department of Computer Science, University of Toronto, Toronto, ON, Canada
2School of Computer Science & Engineering, University of Washington, Seattle, WA, United States
3Department of Human Centered Design & Engineering, University of Washington, Seattle, WA, United States

Corresponding Author:
Alex Mariakakis, PhD
Department of Computer Science
University of Toronto
Bahen Centre for Information Technology, Room 4283
40 St. George Street
Toronto, ON, M5S 2E4
Canada
Phone: 1 416 978 5359
Email: mariakakis@cs.toronto.edu

Abstract

Background: Developers, designers, and researchers use rapid prototyping methods to project the adoption and acceptability of their health intervention technology (HIT) before the technology becomes mature enough to be deployed. Although these methods are useful for gathering feedback that advances the development of HITs, they rarely provide usable evidence that can contribute to our broader understanding of HITs.

Objective: In this research, we aim to develop and demonstrate a variation of vignette testing that supports developers and designers in evaluating early-stage HIT designs while generating usable evidence for the broader research community.

Methods: We proposed a method called health concept surveying for untangling the causal relationships that people develop around conceptual HITs. In health concept surveying, investigators gather reactions to design concepts through a scenario-based survey instrument. As the investigator manipulates characteristics related to their HIT, the survey instrument also measures proximal cognitive factors according to a health behavior change model to project how HIT design decisions may affect the adoption and acceptability of an HIT. Responses to the survey instrument were analyzed using path analysis to untangle the causal effects of these factors on the outcome variables.

Results: We demonstrated health concept surveying in 3 case studies of sensor-based health-screening apps. Our first study (N=54) showed that a wait time incentive could influence more people to go see a dermatologist after a positive test for skin cancer. Our second study (N=54), evaluating a similar application design, showed that although visual explanations of algorithmic decisions could increase participant trust in negative test results, the trust would not have been enough to affect people’s decision-making. Our third study (N=263) showed that people might prioritize test specificity or sensitivity depending on the nature of the medical condition.

Conclusions: Beyond the findings from our 3 case studies, our research uses the framing of the Health Belief Model to elicit and understand the intrinsic and extrinsic factors that may affect the adoption and acceptability of an HIT without having to build a working prototype. We have made our survey instrument publicly available so that others can leverage it for their own investigations.

(JMIR Hum Factors 2022;9(1):e30474) doi:10.2196/30474

KEYWORDS
mobile health; survey instrument; health screening; health belief model; path analysis; user design; health technology; health intervention technology; digital health; mobile phone
Introduction

Overview

There are numerous design decisions beyond the rigor of the information being presented in a health intervention technology (HIT) that can affect how people incorporate the HIT’s guidance into their decision-making [1]. These factors can range from the HIT’s visual appearance [2] and message framing [3,4] to people’s beliefs and psychological traits [5,6]. Late-stage evaluation methods such as A/B field testing and randomized controlled trials are designed to help HIT creators explore the ways in which the aforementioned factors might affect people’s decision-making [7-12]. However, deploying an HIT too early can expose people to numerous risks, such as delays in necessary lifestyle changes, postponed diagnoses, and unwarranted stress. User-centered design also encourages designers to incorporate feedback early and often in their process before reaching these late-stage evaluation methods [13]. Unfortunately, early-stage evaluation and rapid prototyping methods (eg, think-aloud evaluations and paper prototyping) are not as well-suited for eliciting feedback on how people would respond to an HIT’s guidance. Many people assess the credibility of an HIT based on its visual appearance and language [2,14], which may not be fully developed in a low-fidelity prototype. People can also idealize unspecified HIT features to their liking, resulting in a positive but biased evaluation [15]. Even when a prototype is complete, early-stage methods are better suited for identifying which features people prefer but not why they prefer those features or how those features will affect use [16].

In light of these challenges, Klasnja et al [17] called for early-stage evaluation methods that generate usable evidence: “empirical findings about the causal effects of [HITs] and how those effects vary with individual differences, context of use, and system design.” Klasnja et al [17] discussed usable evidence in the context of developers and designers who are creating a new HIT; however, there is also a broader need within the research community to generate findings that lead to guidelines and theories. Identifying usable evidence requires an explicit understanding of the causal mechanisms that affect the reception of an HIT [18], which can only be gained by untangling the effects of HIT design decisions and proximal cognitive factors such as beliefs and attitudes.

As a methodological contribution to HIT design research, we propose health concept surveying, a variation of vignette testing [19,20] that supports the generation of usable evidence. Health concept surveying is centered on a survey instrument that presents target users with a technology concept in a scenario and then measures the potential impact that HIT design decisions may have on 2 distal outcomes [21,22]: (1) adoption of an HIT, which is a person’s intention of using an HIT, and (2) acceptability of an HIT’s suggestions, which is a person’s willingness to conduct the follow-up actions recommended by the HIT.

The survey instrument also measures proximal cognitive factors as defined by a health behavior change framework (eg, the Health Belief Model, HBM [23,24]). The responses to the survey were analyzed using path analysis to surface causal pathways that inform future research on HITs. As health concept surveying relies on design concepts rather than physical prototypes, HIT creators can be selective about which HIT design characteristics they include to prevent study participants from getting distracted by missing or incomplete features.

We demonstrate the efficacy of health concept surveying using 3 case studies to display its utility for multiple stakeholders. The first 2 case studies show how health concept surveying would be beneficial to a developer or designer invested in a particular HIT, whereas the third case study highlights how researchers could use health concept surveying to test a broader hypothesis across multiple HITs. The case studies are centered on sensor-based health-screening apps—smartphone apps that use on-device sensors such as cameras and microphones to identify the presence of medical symptoms—as this domain is emerging in academia and industry alike [25]. The design decisions that are explored in these case studies include (1) the inclusion of an incentive, (2) the inclusion of visual test result explanations, and (3) the trade-off between the true positive rate and true negative rate.

In summary, our research contributes the following:

1. The health concept surveying method, which uses vignette testing to disentangle the effects of design decisions and proximal cognitive factors on the adoption and acceptability of an HIT.
2. Case studies that show how health concept surveying can be used to benefit specific HIT designs while generating usable evidence for the broader community.
3. A more complex case study that shows how health concept surveying can also support more abstract research to directly contribute to our understanding of HITs.

Prior Work

Our research is primarily inspired by a collection of commentaries on behavior change technologies (BCTs) by Klasnja et al and Hekler et al [17,26,27]. BCTs aim to persuade a person to change their habits, whereas HITs can include both health-focused BCTs and technologies that provide a 1-time suggestion for a course of action.

In this thread of research, Klasnja et al [26] first recognized that demonstrating behavior change for early-stage BCTs is often “infeasible as well as unnecessary for a meaningful contribution to HCI research” and instead suggest that researchers strive for “a deep understanding of the how and why of the system use by its target users.” They proposed that researchers can work toward such an understanding by tailoring their evaluation methods to the intervention strategies involved in their HIT (eg, self-monitoring, conditioning, and tunneling [28]), which can require the development of new strategies that balance abstraction with contextual relevance [27]. By leveraging behavioral science theories, Klasnja et al [17] suggested that researchers can not only advance their particular intervention but also generate usable evidence: “empirical findings about the causal effects of BCTs and how those effects vary with individual differences, context of use, and system design.”

Evaluation methods such as factorial designs [7,8], microrandomized trials [9,10], and single-case experimental
designs [11,12] can be used to methodically test hypothesis-driven research; however, these methods are typically considered only after a prototype is sophisticated enough to be put into people’s hands. By using a survey method, health concept surveying allows investigators to include as few or as many details about an HIT as they deem fit. This flexibility of abstraction not only makes the health concept surveying suitable for developers and designers with early-stage HITs but also for researchers as they explore hypotheses around HIT concepts. Health concept surveying also relies on health behavior change frameworks so that researchers can disentangle complicated relationships between factors to generate usable evidence.

**Theory: HBM**

Social psychologists have proposed various frameworks to predict, explain, and change health behaviors in matters related to public and personalized health. These frameworks have been applied to topics ranging from smoking cessation and exercise [29] to vaccination [30] and hearing loss prevention [31]. Health behavior change frameworks typically fall into two categories [32]: social cognition models (eg, theory of planned behavior [33] and HBM [23,24]), which use cognitive factors such as beliefs and attitudes as proximal determinants of behavior; and stage models (eg, transtheoretical model [34]), which describe decisions as a sequence of discrete phases.

Survey instruments for applying health concept surveying could be modeled after any of the aforementioned health behavior change frameworks to specify proximal cognitive factors. In this work, we demonstrate health concept surveying with a survey instrument based on the HBM. Researchers have criticized aspects of the HBM, such as its lack of applicability outside of health-related contexts [35,36] and the inconsistency in how different researchers define its constructs [35,37,38]. Nevertheless, we use the HBM because of its specific focus on health interventions, its applicability to both short-term actions and long-term behaviors, and the potential for its constructs to map to actionable feedback for developers, designers, and researchers. By providing a survey instrument that others can use, we hope to provide standardized questions that mitigate inconsistency.

The HBM posits that a person will undergo an action to improve or maintain their health if the perceived barriers to that particular action are outweighed by the perceived seriousness of the health problem, the perceived susceptibility to that health problem, and the perceived benefits of taking action. All of these constructs are affected by modifying variables, that is, demographic information and psychological characteristics that can explain a person’s decision-making. For instance, someone who is well-educated may understand the benefits of early screening, whereas someone who does not have flexible income may view the cost of a screening examination as burdensome.

Conceptually, the HBM can be summarized using the following equation:

\[
\text{Modifying variables} \times (\text{Seriousness} + \text{Susceptibility} + \text{Benefits} – \text{Barriers}) + \text{Cues to action} = \text{Likelihood of action}
\]

Definitions of the HBM constructs according to Urich [39] are provided in Textbox 1.

**Textbox 1.** The constructs of the Health Belief Model and their definitions.

<table>
<thead>
<tr>
<th>Health Belief Model constructs and definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Perceived seriousness: a person’s subjective assessment of the severity of the health problem and its potential consequences</td>
</tr>
<tr>
<td>• Perceived susceptibility: a person’s subjective assessment of their risk of developing the health problem</td>
</tr>
<tr>
<td>• Perceived benefits: a person’s subjective assessment of the value in taking a certain action</td>
</tr>
<tr>
<td>• Perceived barriers: a person’s subjective assessment of the obstacles to taking a certain action</td>
</tr>
<tr>
<td>• Modifying variables: individual characteristics (demographic and psychosocial) that can affect a person’s perception of a health problem</td>
</tr>
<tr>
<td>• Cues to action: internal or external triggers that prompt a certain action</td>
</tr>
</tbody>
</table>

**Methods**

**Overview**

Health concept surveying is centered on a survey instrument that allows investigators to measure proximal cognitive factors while manipulating HIT features. In this section, we provide details on the contents of the survey instrument, as illustrated in Figure 1. We illustrated this survey instrument with a concept for a sensor-based health-screening app called SkinCheck, which analyzes the appearance of a person’s mole to determine whether it is cancerous. A complete example of the survey instrument used can be found in Multimedia Appendix 1.
Figure 1. The structure of the survey instrument for health concept surveying comprises four stages: (1) preintervention, (2) intervention, (3) postintervention, and (4) end of survey. HBM: Health Belief Model; HIT: health intervention technology.

Survey Design

Preintervention

Our survey instrument starts by presenting respondents with a scenario that describes a cue to action related to the health topic of interest for the HIT. Cues to action can include the emergence of symptoms, promotional advertising, or even direct recommendations or prescriptions from a physician. For our example regarding a sinus infection, our prompt was as follows:

You recently noticed a new mole (beauty mark) on your arm that is oddly colored and misshapen. After looking up information online, you worry that you might be developing skin cancer.

After reading the scenario, the respondent is asked to complete an instructional manipulation check (IMC) [40], where they are asked to select the symptoms that are associated with the described medical condition. In addition to checking that the respondent read the scenario, the IMC forces the respondent to spend extra time reflecting on the scenario.

The respondent is then asked a series of questions related to their initial reactions to the scenario according to the constructs of the HBM: PerceivedSeriousness, PerceivedSusceptibility, PerceivedBenefits, and PerceivedBarriers (Textbox 2). Each construct has a corresponding question except for PerceivedSeriousness, which has 3 questions to account for the various impacts that a health-related issue can have on a person’s life. All responses are recorded on a 7-point scale. The respondent is also asked whether they would take various actions as a series of yes-or-no questions. The respondent is free to take 0, 1, or multiple actions; therefore, we use the variable ActionType to keep track of which action corresponds to each response and ActionTaken to track whether the respondent would take each action. As people can foresee different PerceivedBenefits and PerceivedBarriers for various actions, we also ask the respondent to separately rate those questions for each ActionType.

Textbox 2. The set of questions that are asked in the pre- and postintervention stages of the health concept surveying survey instrument.

Health Belief Model constructs and survey questions

- Perceived seriousness
  - If you had [medical condition] in this scenario, how impactful do you believe it would be on your long-term health?
  - If you had [medical condition] in this scenario, how impactful do you believe it would be on your finances?
  - If you had [medical condition] in this scenario, how impactful do you believe it would be socially and/or professionally?

- Perceived susceptibility
  - How likely do you think you are to have [medical condition] in this scenario?

- Perceived benefits
  - How beneficial do you believe each of these actions would be towards helping you recover from your symptoms?

- Perceived barriers
  - How easy do you think it would be for you to take each of the following actions to help you recover from your symptoms?

- Action taken
  - Given the possibility that you may [have/not have] [medical condition], which of the following actions would you plan to take on the same day as when you discovered your symptoms?
**Intervention**

After the respondents report which actions they would take, they are given information about an HIT that is meant to address the health-related issue described in the scenario. This is where the investigator can choose which details to include about their HIT. Although more details will generally make the HIT concept more concrete and leave less room for uncertainty, the investigator may choose to leave out some information to avoid potential distractions from their primary questions. Our SinusCheck example includes the following text:

*A smartphone app named SkinCheck analyzes a picture of a mole to determine whether or not it is cancerous. To use the app, you are asked to take a picture of the mole so that it is clearly visible. The app guides you through taking a picture so that it can see the mole clearly and at a proper distance.*

SkinCheck comes with your smartphone by default as part of a new mobile health initiative by [Phone Company]. SkinCheck provides text-based and audio-based instructions to help you perform the test. The app also checks that the test was performed correctly. You can repeat the test until the app determines the image to be “valid.” The results of the test are available instantly.

This example includes a high-level description of the app’s source and functionality; however, it does not include any mockups or screenshots of the app itself. Therefore, an investigator could use this example early in their development process to explore how people would feel about the concept of using an app to detect sinus infections without undue influence from the visuals of the app itself, which could be addressed at a later time.

**Postintervention Stage**

After reading the HIT description, the respondent is asked about their interest in using the HIT on a 7-point scale, which we call TechnologyInterest. If the respondent says that they would use the HIT beyond the neutral score, they are taken to pages where they are asked to react to different outcomes in a randomized order. For health-screening apps, our outcomes included positive and negative test results. After each outcome, the respondent is asked to re-evaluate their responses to the questions in Textbox 2. We can determine whether the HIT would have changed the respondent’s plan by comparing ActionTaken across the pre- and postintervention stages. This produces a second outcome variable called ActionChange, indicating whether the HIT had sufficient influence to change a person’s behavioral intent. Similar to ActionTaken, ActionChange is recorded for each ActionType.

Every HBM construct would ideally be evaluated before and after the intervention to examine how perceptions changed as a result of the intervention. However, doing so can significantly increase the survey length when evaluating multiple versions of an HIT. Therefore, an investigator may choose to remove a postintervention question for a particular HBM if they are confident that their design question is unrelated to it. In such cases, the response from the preintervention stage is propagated through the rest of the respondent’s data, as it is assumed to be constant. We use this modification in our third case study as it has 3 manipulated factors and a mixed factorial study design.

**End of Survey**

At the end of the survey instrument, the respondent is asked for information related to ModifyingVariables within the HBM. These questions can capture demographic information (eg, age and access to health care services), psychological properties (eg, risk aversion), or self-assessed expertise in topics related to the HIT (eg, numeracy and familiarity with the medical condition). As the content of the survey itself can provide new information to respondents, some of these questions may be best placed at the beginning of the survey.

**Design Summary**

To summarize, our survey instrument captures two key outcome variables: (1) TechnologyInterest, which measures the likelihood that the respondent would use the app on a 7-point scale, and (2) ActionTaken, which measures the likelihood that the respondent would take action based on the information available to them at that point in the survey. All respondents would answer questions related to each HBM construct, TechnologyInterest and ActionTaken in the preintervention stage. Respondents who express sufficient interest in using the HIT are then shown various potential outcomes of the HIT and asked to reanswer the HBM construct and ActionTaken questions for each one. The responses to ActionTaken in the pre- and postintervention stages are compared for each HIT outcome to form the outcome variable ActionChange. ActionChange is not recorded for respondents who do not express interest in using the HIT as they never reach the postintervention stage. We use TechnologyInterest to project the potential adoption of an HIT, and we use ActionChange to project the potential acceptability of an HIT.

**Analysis**

We analyzed data from our survey instrument using path analysis, a variant of structural equation modeling that discerns the effects of a set of observable variables on a specified outcome via multiple causal pathways [41]. Path analysis revolves around graphical models called path diagrams, which encode hypothesized causal relationships by using nodes to represent measured constructs and directed edges to represent the relationships between them. Running path analysis produces a model in which each edge is assigned a path coefficient and a corresponding $P$ value. We reported standardized path coefficient ($b$), where $b=0.5$ from $X$ to $Y$, suggesting that a 1 SD change in $X$ produces a 0.5 SD change in $Y$.

The result of path analysis is a model in which each edge in the path diagram is assigned a path coefficient and $P$ value. The coefficient is not a correlation coefficient but rather indicates the degree to which one variable influences the other. Chin [42] asserted that meaningful path coefficients have absolute magnitudes $>0.2$. The models themselves can be assessed according to a variety of fit statistics with no agreed-upon standard [43-45]. We reported two fit statistics: comparative fit index (CFI) and standardized root mean square residual (SRMR). CFI compares the model fit against the fit of an
independent model in which the variables are assumed to be uncorrelated, whereas SRMR compares the difference between the residuals of the covariance matrix and the hypothesized covariance model while standardizing for elements with different ranges. Hu and Bentler [46] considered a model fit to be strong when its CFI is ≥ 0.95 and its SRMR is ≤ 0.09. The fit statistics are likely to be poor if the path diagram is insufficient for characterizing the relationship between variables (eg, missing nodes or edges) or if the responses to key variables are heavily biased.

It is possible to analyze the data that are gathered with our survey instrument using techniques such as analysis of variance or generalized linear models; however, separate regressions would be needed for each variable with an inbound edge to capture all the causal pathways in the path diagram. Path analysis makes it easier for investigators to contrast the importance of 2 causal relationships as the entire path diagram is processed at once, and the edge weights are directly comparable. Path analysis also makes it possible to characterize the mediated relationships. In other words, the influence of $X$ on $Z$ via $Y$ can be calculated by multiplying the edge weights from $X$ to $Y$ and from $Y$ to $Z$.

The basic path diagrams used to disentangle the effects that health intervention technology design decisions and user-intrinsic factors have on the measured outcome variables: TechnologyInterest (left) and ActionChange (right). HIT: health intervention technology.

We fit the TechnologyInterest model to the data from all respondents using their ratings for the HBM constructs in the preintervention stage. Models for ActionChange require using data from both the pre- and postintervention stages, therefore limiting the analysis to data from respondents who expressed sufficient interest in using the HIT. Variables such as ActionType do not have causal effects but still produce unique entries in the data set. Rather than including these variables in the path diagrams, they are used as grouping factors for multigroup path analysis, a technique in which a model is fit for each group with assumptions about which attributes the models share. As people can have asymmetric reactions to positive and negative test results, we fit separate ActionChange models in response to positive (ActionChangePositive) or negative (ActionChangeNegative) test results when applicable. In each of these cases, we excluded respondents who would have taken the HIT’s target action in the preintervention stage. For example, respondents who would have taken action in the preintervention stage were excluded from the model because a positive test result would not be needed to convince them to take action.

Figure 2 shows the diagram of our outcome variables. PerceivedSeriousness is a latent variable that combines the responses to its 3 constituent questions. The more nodes that are in the path diagram, the more complicated the model becomes and the more participants that must be recruited to achieve statistical significance. Therefore, we encourage investigators to remove directed edges between 2 variables if they are confident that the variables are unrelated according to their definition or the investigators’ best judgment. For example, we assume that TechnologyInterest is independent of PerceivedBenefits and PerceivedBarriers as those constructs relate to actions that are unrelated to using the HIT itself. HIT design variables and ModifyingVariables should also be added at the investigators’ discretion, with particular focus paid to when they are introduced in the survey instrument. If a design decision affects how the HIT is introduced, the corresponding variable should be added to both path diagrams; however, if the design decision only appears during the intervention stage, the variable should not be included in the TechnologyInterest diagram.

Figure 2. The basic path diagrams used to disentangle the effects that health intervention technology design decisions and user-intrinsic factors have on the measured outcome variables: TechnologyInterest (left) and ActionChange (right). HIT: health intervention technology.

Results

Overview

To demonstrate the flexibility of our method in a series of case studies, we first had to create a variety of prompts for plausible health-related scenarios and sensor-based health-screening apps. We selected three scenarios based on their plausibility and the different reactions we expected them to elicit: (1) a scenario involving pink eye, which represents a common medical condition; (2) a scenario involving skin cancer, which represents a serious medical condition; and (3) a scenario involving halitosis, which represents a stigmatizing medical condition. Multimedia Appendix 2 [47-58] explains the formative study by which these categories and scenarios were selected.

We used these scenarios to generate 3 case studies that highlighted the diverse ways in which health concept surveying can be used. Our first 2 case studies, which are centered around the skin cancer scenario described in the previous section, illustrate how an HIT developer or designer can use health concept surveying to decide whether to include a feature in their
HIT. Our third case study relies on all 3 scenarios to demonstrate how a human–computer interaction (HCI) researcher can use health concept surveying to elicit usable evidence without focusing on a single HIT. We restricted our investigation to a single ActionType (scheduling an appointment) for brevity; however, we featured multigroup path analysis in case study 3 to account for its mixed factorial design and demonstrate the expressivity of our method.

Recruitment
As our case studies were centered on health-screening apps, we recruited participants from the general population without any inclusion or exclusion criteria regarding their experiences with the relevant medical conditions. We sent calls for participation through Facebook, Reddit, and a mailing list within the University of Washington’s Institute of Translational Health Sciences, a center sponsored by the National Institutes of Health’s Clinical and Translational Science for connecting clinicians, patients, and other communities throughout the northwest United States. We excluded respondents who were aged <18 years or did not own a smartphone. Respondents electronically consented before viewing any of the survey materials. Respondents who completed the survey were eligible for a raffle in which 1 in 20 people would win a US $20 Amazon gift card. We used this recruitment strategy for all 3 of our case studies with approval from the University of Washington’s Institutional Review Board (#00003540). Participants were restricted from taking part in multiple case studies to avoid any potential carryover effects or biases (eg, learning and fatigue).

Case Study 1: Incentivizing Clinical Visits

Overview
Our first case study investigated whether the inclusion of a wait time guarantee provides a sufficient incentive for people who would not normally seek medical attention to change their minds and get treatment. We explored this question in the context of our serious medical condition scenario regarding skin cancer. We recruited 54 respondents for this case study, and their demographic information can be found in Table 1.

<p>| Table 1. Demographic information for the people who completed the survey in case study 1 (N=54). |</p>
<table>
<thead>
<tr>
<th>Survey demographics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>6 (11)</td>
</tr>
<tr>
<td>ITHS</td>
<td>48 (89)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>41 (76)</td>
</tr>
<tr>
<td>Male</td>
<td>11 (20)</td>
</tr>
<tr>
<td>Gender variant/nonconforming</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>31 (57)</td>
</tr>
<tr>
<td>25-34</td>
<td>13 (24)</td>
</tr>
<tr>
<td>35-44</td>
<td>7 (13)</td>
</tr>
<tr>
<td>45-54</td>
<td>1 (2)</td>
</tr>
<tr>
<td>55-64</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Smartphone operating system</td>
<td></td>
</tr>
<tr>
<td>iOS</td>
<td>34 (63)</td>
</tr>
<tr>
<td>Android</td>
<td>20 (37)</td>
</tr>
<tr>
<td>Self-reported smartphone experience</td>
<td></td>
</tr>
<tr>
<td>Expert or advanced</td>
<td>32 (59)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>21 (39)</td>
</tr>
<tr>
<td>Novice or beginner</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

*aITHS: Institute of Translational Health Sciences.

Study Design
Figure 3 shows the survey design used in this study. We modified the intervention stage so that respondents were shown 1 of the 2 app descriptions at random. Half of the respondents read the SkinCheck description presented in the Methods section, whereas the other half saw the same description with the addition of the following text to describe a wait time incentive:

Because of their mobile health initiative, [Phone Company] has an exclusive partnership with dermatologists across the country. People who have a questionable mole on their skin according to
SkinCheck are given a promotional code that they can redeem at their local dermatologist to guarantee a wait time no longer than 10 minutes.

As the incentive was intended to make it easier for a person to see a clinician, we only asked respondents about how they would react to a positive test result. The study had a single-factor between-subjects design with the inclusion of an Incentive as the factor of interest. As our lone modifying variable, we asked respondents to rate how quickly they thought they would be able to see their physician as we hypothesized that people who did not have convenient access to a clinician would be more influenced by the incentive. We called this variable Convenience, and it was measured on a 7-point scale. Incentive and Convenience were connected to all major HBM constructs and outcome variables in our path diagrams.

This survey had a completion rate of 83% when we accounted for respondents who ended the survey early, satisfied the exclusion criteria, or did not correctly answer the IMC embedded in the survey. Ignoring 2 cases where respondents took more than an hour-long break while completing the survey, the median survey completion time was 8 minutes.

Figure 3. The survey structure for case study 1. The inclusion of an incentive in the health intervention technology description was randomized across respondents. HBM: Health Belief Model; HIT: health intervention technology.

TechnologyInterest

Most respondents expressed interest in using the SkinCheck app. Of the respondents who completed the survey, 54% (29/54) gave the highest rating possible for TechnologyInterest, 19% (10/54) gave the second-highest rating, 13% (7/54) gave the third-highest rating, and the remaining 15% (8/54) gave ratings that were either neutral or worse. The heavy bias in TechnologyInterest meant that a strong model fit could not be found for this outcome variable (CFI=0.839; SRMR=0.131).

Table 2. Path analysis coefficients for ActionChangePositive in case study 1 (CFI^2=0.951; SRMR^2=0.079).^c

<table>
<thead>
<tr>
<th>Variables</th>
<th>ActionChange</th>
<th>Seriousness</th>
<th>Susceptibility</th>
<th>Benefits</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>AppResult</td>
<td>0.874^d</td>
<td>−0.002</td>
<td>0.636^e</td>
<td>0.024</td>
<td>−0.035</td>
</tr>
<tr>
<td>Incentive</td>
<td>1.138</td>
<td>−0.406</td>
<td>0.275</td>
<td>0.598</td>
<td>−0.361^e</td>
</tr>
<tr>
<td>Convenience</td>
<td>0.128^e</td>
<td>−0.492</td>
<td>0.055</td>
<td>−0.168</td>
<td>−0.384^f</td>
</tr>
<tr>
<td>Seriousness</td>
<td>−0.005</td>
<td>N/A^g</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Susceptibility</td>
<td>0.482^e</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Benefits</td>
<td>0.402^e</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Barriers</td>
<td>−0.791^d</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

^aCFI: comparative fit index.
^bSRMR: standardized root mean square residual.
^cThe columns indicate dependent variables, whereas the rows indicate independent variables.
^dP<.001.
^eP<.05.
^fP<.01.
^gN/A: not applicable.

The model fit had a large positive coefficient from AppResult to ActionChangePositive (b=6.874; P<.001), which was expected because respondents had to see a test result to change their opinion. There was also a strong positive coefficient from AppResult to PerceivedSusceptibility (b=0.636; P<.05), which supported our intuition that a positive test result should increase a person’s perceived likelihood of having skin cancer. ActionChangePositive was heavily influenced by most of the HBM constructs. The strongest influence came from PerceivedBarriers (b=−0.791; P<.001), which was negative as barriers make it more difficult for a person to be able to take action.
Although there were strong coefficients from Incentive to all HBM constructs, the only statistically significant relationship was from Incentive to PerceivedBarriers ($b=-0.361; \ P<.05$). The fact that there is a negative coefficient between the 2 supported our expectation that the incentive would diminish the obstacles that respondents would foresee in the scenarios. Combining this finding with the strong negative coefficient from PerceivedBarriers to ActionChangePositive implies that Incentive had a strong positive effect on ActionChangePositive mediated by PerceivedBarriers. However, the coefficient from Convenience to PerceivedBarriers ($b=-0.384; \ P<.01$) is slightly larger in magnitude than that from Incentive, which indicates that the incentive was somewhat less important than the convenience of getting to a clinician in the first place. Further investigation into our data set revealed that most individuals who decided to take action after seeing a positive test result paired with an incentive gave less than a neutral rating for Convenience; the Convenience ratings for the individuals who were not shown an incentive were more evenly distributed.

**Case Study 2: Presentation of Results**

**Overview**

Our second case study investigated how the presentation of test results may influence a person’s decision-making. We examined whether the inclusion of visuals that explain an algorithm’s decision would engender more trust in an app’s test result. As before, we explored this question in the context of our serious medical condition scenario regarding skin cancer. We recruited 54 respondents for this case study, and their demographic information can be found in Table 3.

<table>
<thead>
<tr>
<th>Survey demographics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>3 (6)</td>
</tr>
<tr>
<td>ITHS$^a$</td>
<td>51 (94)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>45 (83)</td>
</tr>
<tr>
<td>Male</td>
<td>8 (15)</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>34 (63)</td>
</tr>
<tr>
<td>25-34</td>
<td>13 (24)</td>
</tr>
<tr>
<td>35-44</td>
<td>2 (4)</td>
</tr>
<tr>
<td>45-54</td>
<td>2 (4)</td>
</tr>
<tr>
<td>55-64</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Smartphone operating system</strong></td>
<td></td>
</tr>
<tr>
<td>iOS</td>
<td>39 (72)</td>
</tr>
<tr>
<td>Android</td>
<td>15 (28)</td>
</tr>
<tr>
<td><strong>Self-reported smartphone experience</strong></td>
<td></td>
</tr>
<tr>
<td>Expert or advanced</td>
<td>28 (52)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>26 (48)</td>
</tr>
</tbody>
</table>

$^a$ITHS: Institute of Translational Health Sciences.

**Study Design**

Figure 4 shows the survey design used in this study. We modified the postintervention stage so that respondents would be asked to react to both positive and negative test results. Instead of explaining the test result in a paragraph, as in the previous case study, respondents were shown 1 of 2 result screen concepts, illustrated in Figure 5 [47], at random. Both screens were derived from the DermoScreen app by Wadhawan et al [47], which explains diagnostic decisions using the ABCD rule of dermatoscopy [59].

The study had a single-factor between-subjects design with the inclusion of Visuals as the factor of interest. As our lone modifying variable, we asked respondents about their highest level of education as we hypothesized that reading comprehension would affect their understanding of the visualizations; we called this variable Education. Visuals and Education were connected to all major HBM constructs and outcome variables in our path diagrams.

This survey had a completion rate of 82% when we accounted for respondents who ended the survey early, satisfied the exclusion criteria, or did not correctly answer the IMC.
embedded in the survey. Ignoring 1 case when a respondent took more than an hour-long break while completing the survey, the median survey completion time was 9 minutes.

**Figure 4.** The 2 possible interface options that respondents could have been shown in case study 2 when presented with a positive test result: the interface with text descriptions only (left) and the interface with text and visuals to illustrate how the results were obtained (right). The interfaces were primarily inspired by the DermoScreen app by Wadhawan et al [47]. HBM: Health Belief Model; HIT: health intervention technology.

**Figure 5.** The two possible interface options that respondents could have been shown in Case Study 2 when presented with a positive test result: (left) the interface with text descriptions only and (right) the interface with text and visuals to illustrate how the results were obtained. The interfaces were primarily inspired by Wadhawan et al.’s [47] DermoScreen app.

**TechnologyInterest**

Most respondents expressed interest in using the SkinCheck app. Of the respondents who completed the survey, 56% (30/54) gave the highest rating possible for **TechnologyInterest**, 19% (10/54) gave the second-highest rating, 15% (8/54) gave the third-highest rating, and the remaining 11% (6/54) gave ratings that were either neutral or worse. The heavy bias in **TechnologyInterest** meant that a strong model fit could not be found for this outcome variable (CFI=0.874; SRMR=0.096).

**ActionChangePositive**

Across all respondents who expressed sufficient interest in using the app, there were 56% (27/48) of cases when people said that they would not have acted before using the app. After being presented with a positive test result, 78% (21/27) changed their mind: 38% (8/21) were shown visuals and 62% (13/21) were not. The inclusion of explanations clearly had an impact on people’s reaction to the positive test result as the frequency of **ActionChangePositive** was much higher than in the first case study. In fact, there were so few cases when people did not act even after seeing a positive test result that there was not enough data to generate a meaningful model fit (CFI=0.644; SRMR=0.197).

**ActionChangeNegative**

Table 4 shows the causal path coefficients for the **ActionChangeNegative** model fit. Across all respondents who expressed sufficient interest in using the app, there were 52% (25/48) of cases when people said that they would have acted before using the app. After being presented with a negative test result, 48% (12/25) changed their mind: 50% (6/12) were shown visuals and 50% (6/12) were not.

As **ActionChangeNegative** is positive when a person is swayed to not act in the postintervention stage, we expected many of the path coefficients to be negated relative to those observed with **ActionChangePositive** in the first case study. This expectation was confirmed in a couple of instances. First, the negative coefficient from **AppResult** to **PerceivedSusceptibility**
confirmed our intuition that a negative test result should decrease a person’s belief that they had skin cancer in this scenario. Second, the negative coefficient from PerceivedSeriousness to ActionChangeNegative (b=−0.220; P<.05) showed that people who were not as concerned about skin cancer were more likely to change their course of action.

As we hypothesized, including additional information in the form of visuals strengthened respondents’ confidence in their test results. This was reflected in the negative coefficient from Visuals to PerceivedSusceptibility (b=−0.961; P<.01); when shown a negative test result with visuals, respondents were less likely to believe they had skin cancer. However, PerceivedSusceptibility was not influential on ActionChangeNegative (b=−0.056, not significant); therefore, the inclusion of visuals had neither a direct nor indirect effect on a person’s decision to change their action. We also found that Education was not an influential factor for any of the measured constructs or outcome variables.

Table 4. Path analysis coefficients for ActionChangeNegative in case study 2 (CFI=0.961; SRMR=0.078).c

<table>
<thead>
<tr>
<th>Variables</th>
<th>ActionChange</th>
<th>Susceptibility</th>
<th>Seriousness</th>
<th>Benefits</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>AppResult</td>
<td>6.588d</td>
<td>0.000</td>
<td>−0.222e</td>
<td>0.000</td>
<td>−0.342</td>
</tr>
<tr>
<td>Visuels</td>
<td>0.231</td>
<td>0.235</td>
<td>−0.961f</td>
<td>−0.591</td>
<td>−0.610</td>
</tr>
<tr>
<td>Education</td>
<td>0.087</td>
<td>0.086</td>
<td>0.037</td>
<td>0.107</td>
<td>0.055</td>
</tr>
<tr>
<td>Seriousness</td>
<td>−0.220e</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Susceptibility</td>
<td>−0.056</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Benefits</td>
<td>−0.233g</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Barriers</td>
<td>0.223</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aCFI: comparative fit index.

bSRMR: standardized root mean square residual.

cThe columns indicate dependent variables, whereas the rows indicate independent variables.

dP<.001.

eP<.05.

fP<.01.

gN/A: not applicable.

Case Study 3: Accuracy

Overview

In our third and final case study, we explored the trade-off between false positives and false negatives across medical conditions of varying concern and severity. We leveraged all three of our scenarios (common, serious, and stigmatizing) in a mixed factorial study design, thus necessitating more participants. In total, 263 respondents completed the survey from start to finish, and their demographic information can be found in Table 5.
Table 5. Demographic information for the people who completed the survey in case study 3 (N=263).

<table>
<thead>
<tr>
<th>Survey demographics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>16 (6.1)</td>
</tr>
<tr>
<td>ITHS(^a)</td>
<td>240 (91.3)</td>
</tr>
<tr>
<td>Reddit</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (1.5)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>202 (76.8)</td>
</tr>
<tr>
<td>Male</td>
<td>45 (17.1)</td>
</tr>
<tr>
<td>Transgender male</td>
<td>5 (1.9)</td>
</tr>
<tr>
<td>Gender variant/nonconforming</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>Self-identify</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>145 (55.1)</td>
</tr>
<tr>
<td>25-34</td>
<td>84 (32)</td>
</tr>
<tr>
<td>35-44</td>
<td>17 (6.5)</td>
</tr>
<tr>
<td>45-54</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>55-64</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>≥65</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td><strong>Smartphone operating system</strong></td>
<td></td>
</tr>
<tr>
<td>iOS</td>
<td>170 (64.6)</td>
</tr>
<tr>
<td>Android</td>
<td>93 (35.4)</td>
</tr>
<tr>
<td><strong>Self-reported smartphone experience</strong></td>
<td></td>
</tr>
<tr>
<td>Expert or advanced</td>
<td>146 (55.5)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>115 (43.7)</td>
</tr>
<tr>
<td>Novice or beginner</td>
<td>2 (0.8)</td>
</tr>
</tbody>
</table>

\(^a\)ITHS: Institute of Translational Health Sciences.

**Study Design**

Figure 6 shows the survey design for this study, which required changes in both the intervention and postintervention stages. The app descriptions included information about their classification sensitivity and specificity; sensitivity refers to the proportion of people who are correctly identified as having the medical condition out of all those who have it, whereas specificity refers to the proportion of people who are correctly identified as not having the medical condition out of all those who do not have it. Because the general public is more adept at reasoning about counts than fractional quantities [60], the sensitivity and specificity rates were presented with counts and icon arrays. An example of the accompanying text is provided as follows:

_out of every 100 people who have a sinus infection, SinusCheck correctly told 65 people that they had a sinus infection._

_out of every 100 people who do not have a sinus infection, SinusCheck correctly told 80 people that they did not have a sinus infection._

The survey was used in a 3×3×3 mixed factorial study design. Each respondent read all three scenarios—pink eye (common), skin cancer (serious), and halitosis (stigmatizing)—making ConditionType a within-subjects factor. The presentation order of the scenarios was counterbalanced across all subjects. A total of 3 equally spaced levels of sensitivity and specificity were investigated—65%, 80%, and 95%—producing 9 possible combinations that described the overall accuracy of the apps. Each app for each respondent was assigned 1 of the 9 combinations at random, making Sensitivity and Specificity between-subjects factors. Although there is an inherent trade-off

https://humanfactors.jmir.org/2022/1/e30474

JMIR Hum Factors 2022 | vol. 9 | iss. 1 | e30474 | p.97
(page number not for citation purposes)
between sensitivity and specificity when the underlying classification algorithm is fixed, we treated them as independent variables in our study design and analyses. As respondents had to go through multiple scenarios, we shortened the survey by only remeasuring PerceivedSusceptibility and ActionTaken during the postintervention stage. The other major HBM constructs were not remeasured as we assumed that they should not be influenced by app accuracy. As such, Sensitivity and Specificity were connected to PerceivedSusceptibility and the outcome variables in our path diagrams, and ConditionType was used as the grouping variable for multigroup path analysis. This survey had a completion rate of 73% when we accounted for respondents who ended the survey early, satisfied the exclusion criteria, or did not correctly answer the IMCs embedded in the survey. Ignoring 14 cases when respondents took more than an hour-long break while completing the survey, the median survey completion time was 16 minutes.

**Figure 6.** The survey structure for case study 3. Respondents were shown 3 different health intervention technologies (HITs)—1 for each ConditionType. The 3 HITs either had the same sensitivity and varied in specificity or had the same specificity and varied in sensitivity. Respondents were asked to react to positive and negative app results in a randomized order. Only PerceivedSusceptibility and ActionChange were remeasured in the postintervention stages to shorten the survey length. HBM: Health Belief Model; HIT: health intervention technology.

**TechnologyInterest**

Table 6 shows the causal path coefficients for the TechnologyInterest model fit. The path coefficients from Sensitivity and Specificity to TechnologyInterest were sizable and positive across all scenarios, confirming that higher accuracy made the apps more attractive. In fact, the effect was so strong that those coefficients were larger and more statistically significant than those from the HBM constructs. This suggests that respondents were willing to use these apps regardless of their perception of the medical conditions’ threat as long as they knew that the app was accurate.

<table>
<thead>
<tr>
<th>Variables</th>
<th>TechnologyInterest</th>
<th>Common</th>
<th>Serious</th>
<th>Stigmatizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susceptibility</td>
<td>0.206^d</td>
<td>0.120</td>
<td>0.120^e</td>
<td>0.104</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.416^f</td>
<td>0.357^f</td>
<td>0.268^d</td>
<td>0.292^f</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.461^f</td>
<td>0.300^d</td>
<td>0.129</td>
<td></td>
</tr>
</tbody>
</table>

^aCFI: comparative fit index.
^bSRMR: standardized root mean square residual.
^cThe columns indicate dependent variables, whereas the rows indicate independent variables.
^dP<.01.
^eP<.05.
^fP<.001.

Overall accuracy was most valued for the common condition (Sensitivity: b=0.416, P<.001; Specificity: b=0.461, P<.001), followed by the serious (Sensitivity: b=0.357, P<.001; Specificity: b=0.300, P<.01) and stigmatizing (Sensitivity: b=0.268, P<.01; Specificity: b=0.292, P<.001) conditions. Respondents preferred apps with higher accuracy; however, they attributed more importance to sensitivity or specificity depending on the scenario; they placed more importance on specificity for the common and stigmatizing conditions, whereas they placed more importance on sensitivity for the serious condition. Sensitivity and specificity were treated independently in our analysis; therefore, these results do not account for the fact that improving one metric often requires compromising the other during the development of the classification model.
Nevertheless, this result suggests that respondents had an inherent knowledge about the notion of prevalence and how it relates to diagnostic decision-making. Common and stigmatizing conditions are typically prevalent; therefore, prioritizing specificity may indicate that respondents were eager to use an app’s test result to rule out having the condition. Serious conditions are often less prevalent; therefore, prioritizing sensitivity may indicate that respondents were eager to rule in having the condition.

### ActionChangePositive

Table 7 shows the causal path coefficients for the ActionChangePositive model fit. Across all respondents who expressed sufficient interest in using any of the 3 apps, there were 56.5% (359/635) of cases when respondents said that they would not have taken action before using the app. After being presented with a positive test result, 46.2% (166/359) changed their mind: 28.3% (47/166) in the common scenario, 42.2% (70/166) in the serious scenario, and 29.5% (49/166) in the stigmatizing scenario.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Common</th>
<th>Stigmatizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stigmatizing</td>
<td>ActionChange Positive</td>
<td>Susceptibility</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.426</td>
<td>0.474</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.283</td>
<td>0.204</td>
</tr>
<tr>
<td>Seriousness</td>
<td>0.398</td>
<td>0.168</td>
</tr>
<tr>
<td>Susceptibility</td>
<td>N/A</td>
<td>0.148</td>
</tr>
<tr>
<td>Benefits</td>
<td>0.226</td>
<td>N/A</td>
</tr>
<tr>
<td>Barriers</td>
<td>0.014</td>
<td>N/A</td>
</tr>
</tbody>
</table>

| Stigmatizing | ActionChange Positive | Susceptibility |
| Sensitivity | 0.474 | 0.283 |
| Specificity | 0.204 | N/A |
| Seriousness | 0.168 | 0.014 |
| Susceptibility | N/A | 0.148 |
| Benefits | N/A | 0.148 |
| Barriers | N/A | 0.014 |

Although there were large positive coefficients from AppResult to ActionChangePositive and PerceivedSusceptibility across all scenarios, the magnitude and significance of those coefficients varied across the medical conditions. The coefficient from AppResult to PerceivedSusceptibility for the Serious condition (b=1.518; P<.001) was 3 times as large and more significant than the corresponding coefficients for the common (b=0.398; P<.001) and stigmatizing (b=0.474; P<.01) conditions. Again, this result suggests that respondents may have been eager to use the positive test result from an app to rule in having a serious condition.

### ActionChangeNegative

Table 8 shows the causal path coefficients for the ActionChangeNegative model fit. We note that this model had borderline significance according to our fit statistics, satisfying the threshold for SRMR but not for CFI. Across all respondents who expressed sufficient interest in using any of the 3 apps, there were 41.9% (266/635) of cases when respondents said that they would not have taken action before using the app. After being presented with a negative test result, 51.9% (138/266) changed their minds: 39.1% (54/138) in the common scenario, 37.7% (52/138) in the serious scenario, and 23.2% (32/138) in the stigmatizing scenario.
As with the model fit for ActionChangePositive, there were statistically significant coefficients from AppResult to PerceivedSusceptibility and ActionChangeNegative; however, their magnitude varied across ConditionType. The coefficients from AppResult to PerceivedSusceptibility for the common (b=−1.999; P<.001) and stigmatizing (b=−2.191; P<.001) conditions were nearly double the corresponding coefficient for the serious condition (b=−0.970; P<.001), suggesting that respondents may have been eager to use the negative test result from those apps to rule out having those conditions.

In the serious condition scenario, significant negative coefficients were found from specificity to PerceivedSusceptibility (b=−0.185; P<.001) and PerceivedSusceptibility to ActionChangeNegative (b=−0.358; P<.001). This combination of results implies that respondents were more likely to be convinced to change their course of action after seeing a negative test result when the app had a higher specificity. Specificity did not have a significant effect on PerceivedSusceptibility in either the common (b=0.094, not significant) or the stigmatizing (b=−0.119, not significant) conditions, indicating that respondents were equally willing to accept a negative test result across the presented specificity rates in those scenarios. Sensitivity did not have a statistically significant effect on either ActionChangeNegative or PerceivedSusceptibility for any of the scenarios, which mirrors the earlier findings with respect to specificity and positive test results.

### Discussion

#### Principal Findings

We sought to develop a low-burden method for projecting the adoption and acceptability of an HIT, given different design variations. Our contribution toward this goal—the health concept surveying method—supports HIT investigators in advancing their own HITs while generating usable evidence for the broader research community. Our 3 case studies highlight the different types of actionable feedback and usable evidence that can be elicited using our survey instrument without deploying a working HIT prototype.

Our first case study showed that a wait time incentive might support some individuals in overcoming barriers that could prevent them from visiting a dermatologist. However, many participants said that they would be persuaded to act without an incentive. This result suggests that HIT developers in this scenario may want to consider additional messaging that targets other facets of the HBM, such as the perceived susceptibility people have to skin cancer or the perceived benefits of seeking a second opinion. We also found that access to convenient health care was an important factor in people’s decision-making; therefore, developers in this scenario may want to examine whether this is an important issue to address for their target audience.

Our second case study showed that SkinCheck’s baseline explanation could be convincing enough to sway a person to visit a clinician when they received a positive test result. The inclusion of visuals increased individuals’ trust in negative test results; however, this was not enough to significantly affect people’s decision-making. In fact, we found that the main driving factor for people who decided not to act after seeing a negative test result was the perceived seriousness of skin cancer. This presents an interesting challenge for HIT designers. Lowering a person’s concern about the severity of a medical condition could have major consequences, including the fact that they may ignore a positive test result later on because of their newfound understanding of the condition. Instead, HIT designers in this scenario may want to consider using a language that diminishes a person’s short-term concerns but encourages repeated testing in the near future.

Our third case study suggests that researchers may want to consider the trade-off between sensitivity and specificity in the context of their target medical condition. Kay et al [61] elicited...
similar findings through a survey instrument they created to understand the acceptability of precision and recall across various sensor-based technologies. In an example involving a home alarm system, they showed that participants were more willing to accept false alarms when the system had a benign intervention (e.g., contacting the homeowners via SMS text message) than when the system had an intrusive intervention (e.g., automatically alerting the police). To improve the user experience that people have with a classifier-based application, HIT developers may consider adjusting the final decision threshold of their classifier to minimize errors that people are more prone to believe. However, doing so may serve as an expedient solution to the greater challenge of helping ordinary people with Bayesian reasoning.

Other Design Decisions for Exploration
We explored the influence of 3 different design choices on outcomes relevant to HITs (incentives, results presentation, and accuracy trade-offs); however, there are many others that would be interesting to explore in future work. One of those factors would be the HIT’s price. When we first piloted our studies, we stated that the apps could be purchased on app stores for US $0.99. We selected such a low cost as we were worried that a free app would appear illegitimate; however, an expensive app would diminish interest to the point that we would not receive feedback from respondents. However, some of the respondents in our pilot study felt that a US $0.99 app appeared less legitimate than a free app and cheap, so we instead crafted scenarios in which the app was already included on the respondents’ phones. The economics research community has debated the relationship between price and perceived product quality; some researchers argue that there is generally a positive correlation between price and quality [62], whereas others argue that the 2 are only correlated under contrived scenarios [63].

Another factor that influences the perceived quality of technology is endorsements [64]. App stores, smartphone manufacturers, special interest groups, and physicians can all endorse technologies, serving as a seal of approval that may imbue an HIT with legitimacy. A limitation of our survey instrument is that it is difficult to convey an endorsement to respondents without explicitly drawing the respondents’ attention to it. Endorsements can appear in many places—commercials, supplemental materials, or websites—that may not be as conspicuous as mentioning would be done in the survey. Determining a more natural way of introducing endorsements within health concept surveying could be a potential avenue for future work.

Alternative and Complementary Approaches
Health concept surveying is one of many early-stage quantitative research methods that developers and designers can use to further their understanding of HITs. Conjoint analysis and discrete choice experiments elicit preferences by asking participants to pick between options with 1 or many feature variations in a head-to-head comparison [65]. Another relevant technique is judgment analysis [66,67], where feature preferences are gathered by comparing the decisions that participants make in hypothetical scenarios against a predefined oracle or reference group. All of these methods have been used to investigate people’s decision-making in the health domain [68-70]; however, health concept surveying has the advantage of being designed so that investigators can project both the adoption of an HIT and the acceptability of an HIT’s suggestions. By accounting for intrinsic and extrinsic factors that can influence these distal outcomes, health concept surveying is able to elicit usable evidence that HIT developers and designers can apply to their own HITs.

We view health concept surveying as being complementary to qualitative research methods such as focus group interviews, which give participants the chance to verbalize their thoughts and decision-making in a richer way than what can be gathered through a survey. That said, health concept surveying is far more efficient to scale. Focus groups must be run with 5 to 10 participants at a time, and investigators must often conduct multiple sessions to reach diverse populations or gather feedback on new design iterations. Each new session incurs an additional time investment for both the interviews and the qualitative analyses, making focus groups difficult to scale as an HIT evolves. In addition, focus groups have known confounds such as group-think or dominance by 1 or 2 individuals, even in light of techniques to mitigate these confounds [71]. With health concept surveying, adding more participants simply requires distributing the survey to more people and then rerunning the same analysis code as before, imposing no additional burden beyond what is required for recruitment. Health concept surveying also helps investigators systematically analyze the influence of all the variables involved in people’s decision-making, which can otherwise be difficult for participants to articulate and for investigators to translate into usable evidence. We hope that our work inspires HCI researchers to explore how people can incorporate psychological frameworks into other evaluation techniques.

Limitations
Several psychological frameworks for explaining behavior rely on the belief that intention is a strong predictor of behavior. The correlation between intention and behavior has been supported by research on health-related topics such as dieting [72], physical activity [73,74], and weight loss [75]. Nevertheless, people’s behavioral intentions or expected actions do not always lead to completing the action because of the emergence of unforeseen barriers or changing beliefs over time. Psychologists have called this phenomenon the intention–behavior gap [76,77]. This potential disconnect exists in most early-stage evaluation methods; however, the gap may be particularly relevant to health concept surveying as intention in scenario-based study designs may not translate to real-world actions, and there are no consequences to hypothetical decisions. Despite these shortcomings, there are steps that HIT investigators can take to engender more confidence in their survey responses. We recruited respondents from the public; however, developers and designers who are creating an HIT for a specific audience may want to recruit participants who are either in an at-risk demographic or actively seeking solutions in the HIT’s target domain. As realism is an important mediator in the intention–behavior gap, we also suggest that investigators craft their scenarios with the help of domain experts to make the scenarios as realistic as possible. Investigators could even add
questions to their surveys that measure the degree to which respondents resonate with their scenarios; such measures could be used to either filter responses or create an additional modifying variable in the analyses.

HIT investigators may also want to consider focusing on short-term actions rather than long-term goals (eg, *I intend to eat more vegetables for dinner today* vs *I intend to lose 10 pounds this month*) when querying how a person would respond to an HIT; intention is believed to be a weaker predictor for long-term goals as completing them requires more self-efficacy and coordination to complete [76]. Finally, the health action process approach of Schwarzer [78] separates preintentional motivation and postintentional volition when measuring the likelihood of action; therefore, doing the same in health concept surveying may be beneficial.

To ensure that we were collecting meaningful responses, we also had to create plausible scenarios. We validated the scenarios used in our work through a pilot study using an abridged version of our survey instrument. Researchers who are investigating high-level questions as we did in our third case study would want to repeat this procedure; however, an HIT developer interested in advancing a particular HIT design while generating usable evidence may only need to assess scenario plausibility. We used a single question that explicitly asked respondents how plausible they believed a scenario to be; however, future researchers may want to investigate the nuances of plausibility through multiple questions. A person may believe a scenario is plausible as the health issue in question is common for their demographic or to people who engage in similar behaviors, or they may believe it is plausible because they do not have enough knowledge about the issue to know better. Researchers interested in examining HIT design decisions across multiple scenarios may also want to consider making their scenarios publicly available for future use. Sharing a common set of prevalidated scenarios would standardize the context of findings related to the same topic (eg, physical activity, step counting, and exercise).

**Conclusions**

As more HITs transition from research to practice, it is important for HCI researchers to examine how those technologies will be received by the general population. Although one-off user studies provide actionable feedback for a specific HIT, they rarely provide insights that benefit other HIT creators. Our method, health concept surveying, attempts to strike a balance between actionable feedback and usable evidence. Using the HBM, health concept surveying disentangles proximal cognitive factors from HIT design decisions to explain *how* and *why* certain features are preferred. We used health concept surveying in 3 case studies to demonstrate the range of questions it can support and discussed the implications of the findings in each case. We hope that researchers will continue using health concept surveying in the future to better our understanding of HITs and accelerate their development.

**Acknowledgments**

This project was supported by funding from the National Science Foundation (NSF IIS-1813675, IIS-1553167) and the National Institutes of Health (NIH R21 DK117431 and NIH R01 LM012810). The authors would like to thank Wanda Pratt for her guidance during the early stages of this research. The authors would also like to thank Gary Hsieh for his feedback on the framing of this work.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
An example of the health concept surveying survey instrument was applied to a design concept called SkinCheck.

[DOCX File, 75 KB - humanfactors_v9i1e30474_app1.docx ]

Multimedia Appendix 2
A preliminary study was conducted to select the scenarios and design concepts that appear in this paper.

[DOCX File, 459 KB - humanfactors_v9i1e30474_app2.docx ]

**References**


78. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. 

Abbreviations

- **BCT**: behavior change technology
- **CFI**: comparative fit index
- **HBM**: Health Belief Model
- **HCI**: human–computer interaction
- **HIT**: health intervention technology
- **IMC**: instructional manipulation check
- **SRMR**: standardized root mean square residual

©Alex Mariakakis, Ravi Karkar, Shwetak N Patel, Julie A Kientz, James Fogarty, Sean A Munson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 03.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Purpose Formulation, Coalition Building, and Evidence Use in Public–Academic Partnerships: Web-Based Survey Study

Christina D Kang-Yi1,2*, PhD; Amy Page3*, DSW

1Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
2Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States
3Graduate School of Social Service, Fordham University, New York City, NY, United States
*all authors contributed equally

Corresponding Author:
Christina D Kang-Yi, PhD
Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania
3535 Market Street, 3rd Floor
Philadelphia, PA, 19104
United States
Phone: 1 215 746 6715
Email: christina.kang-yi@pennmedicine.upenn.edu

Abstract

Background: Partnerships between academic institutions and public care agencies (public–academic partnerships [PAPs]) can promote effective policy making and care delivery. Public care agencies are often engaged in PAPs for evidence-informed policy making in health care. Previous research has reported essential partnership contextual factors and mechanisms that promote evidence-based policy making and practice in health care. However, the studies have not yet informed whether public care agency leaders’ and academic researchers’ perceptions of partnership purpose formulation and coalition building evolve through the PAP life cycle and whether public care agency leaders’ use of research evidence differs through life cycle stages.

Objective: This exploratory study aims to focus on PAPs designed to improve youth mental health and well-being outcomes. This study also aims to identify public care agency leaders’ and academic researchers’ perceptions of PAP purpose formulation (structure, goals, primary function, and agenda-setting process) and coalition building (mutual benefits, trust, convener’s role, member role clarity, and conflict management) by PAP life cycle stage and examine whether public care agency leaders’ use of research evidence differs according to the perception of PAP purpose formulation and coalition building through the PAP life cycle.

Methods: A web-based survey of PAP experience was conducted by recruiting academic researchers (n=40) and public care agency leaders (n=26) who were engaged in PAPs for the past 10 years. Public care agency leaders additionally participated in the survey of the Structured Interview for Evidence Use scale (n=48).

Results: Most public care agency leaders and academic researchers in PAPs formed, matured, and sustained perceived their PAP as having purpose formulation context well aligned with their organizational purpose formulation context, pursuing mutual benefits, having leadership representation and role clarity, having a higher level of trust, and knowing how to handle conflicts. Most PAPs across all life cycle stages crystallized another issue to focus, but not all PAPs with issue crystallization had purpose reformulation. Public care agency leaders who trusted academic researchers in their PAP had greater use of research evidence. Public care agency leaders in PAPs that had gone through new issue crystallization also showed greater use of research evidence compared with those that had not.

Conclusions: To promote public care agency leaders’ use of research evidence, focusing on developing trusting partnerships and continuously crystallizing PAP issues are important.

International Registered Report Identifier (IRRID): RR2-10.2196/14382

(JMIR Hum Factors 2022;9(1):e29288) doi:10.2196/29288

KEYWORDS

use of research evidence; public care policy; public–academic partnership; partnership purpose formulation; partnership coalition building; youth mental health and well-being
Introduction

Background

Partnerships between academic institutions and public care agencies (public–academic partnerships [PAPs]) can promote effective public policy making and care delivery. For example, local US public health departments that formally partner with academic institutions are more likely than those not engaged in partnership with academic institutes to make evidence-based policy making and implement evidence-based interventions in health care delivery [11]. Previous studies have demonstrated the important role of PAPs in training service providers [3-5], supporting the implementation of promising evidence-based practices [3-7], and conducting systems evaluation that inform policy development and program planning [2,4]. Such partnerships have effectively responded to the need for additional, more diverse, and more inclusive mental health and child welfare services [2,3,8-10].

Although previous studies have demonstrated the positive impact of PAPs on youth mental health and well-being outcomes, few empirical studies have examined whether and how PAP contexts and mechanisms evolve through the PAP life cycle and which PAP contexts and mechanisms foster public agency leaders’ use of research evidence to improve youth mental health and well-being [11]. Public mental health and child welfare agencies are expected to increase the use of evidence-based care to improve mental health and well-being of vulnerable youth [12–16]. Many public care agencies partner with academic researchers to meet these expectations. Considering the multifaceted nature of public mental health and child welfare systems in the United States [17-20], it is important to develop a better understanding of the context and mechanisms that promote successful PAPs and evidence use by policy makers to improve youth outcomes.

This study has 3 aims. First, we describe a new integrated framework to understand PAP development through the PAP life cycle and potential PAP contexts and mechanisms that foster public care agency leaders’ use of research evidence. Second, we summarize the literature to provide empirical support for the integrated framework, focusing on the contexts and mechanisms of PAP purpose formulation and coalition building. Third, we report our study that comprehensively explored the relationship between PAP purpose formulation and coalition building and public care agency leaders’ use of research evidence by PAP life cycle stages of formed, matured, sustained, declining, and terminated.

Key PAP Process

Although research on individual components of the partnership process has revealed important information about factors that support successful partnerships, the literature has yet to bring these components together into an integrated framework [11]. Such a framework would offer a way to examine the totality of PAPs, including the contexts in which they initiate and mature, the mechanisms that propel them forward, and the outcomes that they define and achieve at various stages in relation to public care agency leaders’ use of research evidence. The integrated framework by Kang-Yi [11] introduces concrete components of partnership purpose formulation and coalition building as the key contexts and mechanisms of PAPs that lead to policy makers’ use of research evidence. The framework consists of three theoretical perspectives: the social partnerships perspective [21,22], the organizational life cycle perspective [23-26], and the realist evaluation perspective [27,28].

On the basis of the social partnerships and organizational life cycle perspectives [21,25], the integrated framework posits that PAPs that continuously reformulate partnership purposes and build coalitions are likely to successfully evolve through life cycle stages of being formed, matured, and sustained. Public care agency leaders in successful PAPs (being matured or sustained compared with being just formed, declining, or terminated) are more likely to use research evidence. According to the social partnerships perspective by Waddock [21], purpose formulation processes include identifying clear goals and the primary function of partnership, creating a partnership structure, and setting partnership agenda. Coalition building processes include pursuing mutual benefits for each partner, building trust among partners, solidifying the convener’s role, clarifying the roles of all parties, and managing conflict [21]. The realist evaluation perspective provides a methodology for configuring contexts, mechanisms, and outcomes to examine the interplay of partnership purpose formulation, coalition building, and public care agency leaders’ use of research evidence in each PAP life cycle stage and overall evolvement of PAP [27-29]. The integrated framework emphasizes continuous purpose formulation and coalition building to adjust to changing partnership environment, sustain PAP, and promote public care agency leaders’ use of research evidence.

Purpose Formulation

Agenda-Setting

One key ingredient in successful PAPs is the development of a clear purpose formulation among partners. Focusing on the needs of policy makers [2,6,30–32] and having public care agency representatives who are also skilled researchers driving the agenda-setting process are important [2].

Goals

Setting clear goals for a PAP is an important aspect of achieving and measuring success. Clear PAP goals have the power to keep partners focused on working toward positive outcomes [10]. PAPs in which goals are aligned with the goals of each partnering entity can contribute to the success and longevity of those PAPs [4-6]. PAPs with clear goals can promote the use of evidence by policy makers [30].

Primary Function

PAPs can play diverse primary functions, including generating knowledge related to the development and implementation of evidence-based policy making and practices, generalizing practices to a larger population, disseminating knowledge related to the implementation of evidence-based practices, and offering technical assistance, such as professional training and program evaluation in improving service quality and outcomes [11]. Given that public care agencies and academic institutes pursue diverse missions and primary functions, alignment in primary
functions between a public care agency and an academic research institute can influence PAP sustainability and use of evidence by policy makers.

Structure
A partnership structure involves shaping governance processes, agreements around dissemination of findings, data sharing, business arrangements, ethics approvals, determining partnership mission, and general coordination among the partners [5,7,31-34]. The degree and quality of formalized structure shapes the extent of PAP success [5,32,35].

Partnership Coalition Building
Previous studies have shown that coalition building, including mutual benefits and trust, plays a critical role in successful partnerships and in promoting the use of evidence by policy makers. The key dimensions of partnership coalition building include mutual benefits, trust, convener’s role, member role clarity, and conflict management.

Mutual Benefits
Successful PAPs are found to pursue mutual benefits, such as having specific agreement that ensures strategic advantages for both parties, smoother facilitation of contracts, financial incentives for the university, conducting actionable research, offering innovative ideas, improving the quality of services, offering researchers the benefit of evaluating a new theoretical model, and facilitating knowledge translation to direct practice [10,30,34-36]. Although PAPs offer a range of mutual benefits, they are not without risk. For example, time, effort, and cost of work are costs for all parties involved [36]. Risks specific to researchers include opportunity costs of spending time on projects that may not lead to publications and the potential negative impact of a changing political environment [34,36]. Risks for policy makers include spending social capital to justify engagement in the PAP, working with researchers who might not appreciate the complexity involved in the PAP work, the potential that research outcomes might not be practical, and the unknown impact the partnership may have on the organization [34,36].

Trust
Trust is another vital component of partnerships’ success. Trust plays a key role in the sustainability of partnerships, leading to continued work, additional projects, and system-level changes. Trusting relationships among partners also support PAPs in weathering leadership changes, particularly when work has become integral to the functioning of an agency, promoting more efficient and purposeful engagement of policy makers in the research process [31,36]. Trust among partners may also facilitate the use of evidence by policy makers. For example, some PAPs appoint personnel specifically to serve as relationship cultivators and to seek input into research questions to be explored by PAPs [30].

Convener’s Role
PAPs need conveners to bring partners together into a partnership formation. Previous studies have documented the importance of such a role in bringing partners together in long-standing relationships within both organizations, identifying problem areas and developing initiatives in response, maintaining the necessary structure of PAP to disseminate the information generated by the partnership, and promoting the use of research evidence by policy makers [33,35,37]. Individuals who possess knowledge spanning both research and policy realms can support translating knowledge into the policy process [35].

Role Clarity
Clear delineation of roles among partners related to developing research questions and methodology as well as the eventual dissemination of the findings is important for successful PAPs [30,33]. In addition, clear communication between partners about how decisions are to be made and whether researchers can provide policy recommendations is critical [10], as these decisions can make a difference in informing policy makers and promoting the use of research evidence among policy makers. Partnerships that are slow in building comprehensive leadership teams and having members who are unsure of their roles can delay the generation of useful evidence for policy.

Conflict Management
Conflict is not unusual in the life of a partnership. Disagreement over project aims and funding [30] and other partnership processes, such as agenda-setting and contracting, can increase. Effective conflict management skills are important in building successful PAPs that lead to the use of research evidence in policy making.

This Study: Web-Based Survey of PAPs
Our study aims to focus on PAPs designed to improve youth mental health and well-being. This study also aims to identify whether contexts and mechanisms of PAP partnership purpose formulation (structure, goals, and primary function as contexts and agenda-setting process as mechanism) and coalition building (convener’s role, leadership representation, role clarity, and conflict management as contexts and mutual benefits and trust as mechanisms) evolve through PAP life cycle stages (formed, matured, sustained, declining, and terminated). We also examined whether public care agency leaders’ use of research evidence differs according to their perception of the PAP life cycle stage, purpose formulation, and coalition building. Research evidence was defined as relevant conceptual frameworks or reviews and empirical findings from systematic qualitative, quantitative, or mixed research methods projects [38]. The study was approved by the institutional review board of the University of Pennsylvania (see Kang-Yi [11] for the published study protocol).

Methods
Sampling and Participant Recruitment
A web-based survey of PAP partnership experience and use of research evidence was conducted by recruiting academic researchers and public care agency leaders who were engaged in PAPs. See Page et al [39] for a detailed discussion of the approach used to identify PAP researchers and public care agency leaders. To recruit public care agency leaders and academic researchers who were engaged in PAPs, we identified
PAPs through two primary methods: a web-based search of peer-reviewed journals and Google for key terms related to youth-focused PAPs and national and local meetings of professionals and researchers in the fields of mental health and child welfare. A total of 87 PAPs were identified, which met the following criteria: formed on a project, program, or intervention basis or as a consortium; aimed to improve mental health and well-being outcomes for youth aged 12-25 years; and comprised at least one or more state or local county mental health and child welfare agencies and one or more academic researchers. PAPs focused on youth outside the United States or established outside the United States and PAPs terminated within 10 years before the study initiation in 2017 were excluded. Of the 87 PAPs identified, we reached out to at least one public care agency leader in 67 PAPs and at least one academic researcher in 83 PAPs.

Once we identified researchers and public care agency leaders, we emailed them a link to the web-based survey along with introductory information about the nature of the study. A US $35 gift card was offered for full completion of a survey. Respondents were informed that the link was unique to them and asked not to share it with others. Data were collected from March 2019 to February 2020. The survey was tested for usability and accuracy by the research team and a small number of colleagues before being shared with potential respondents. In addition, the Checklist for Reporting Results of Internet E-Surveys [40] was used to report the survey as needed.

Survey Measures
To respond to the survey questionnaire, the participants were asked to focus on the latest PAP or one of the PAPs for the past decade if they were not engaged in PAP at the time of the survey. The Structured Interview for Evidence Use (SIEU) [41] was used to identify public care agency leaders’ engagement level of research evidence, which refers to the frequency of using various types of sources for research evidence; public care agency leaders’ ratings of the importance of evaluating the validity, reliability, and relevance of research evidence; and various factors leading public care agency leaders to use or ignore research evidence in deciding to adopt a new program or intervention. The SIEU was developed based on the posited that research use is driven by context and social relationships [41]. Thus, SIEU as a tool reflects the integrated conceptual framework being tested in this study. SIEU includes input, process, and output scales. The input scale (20 items) assesses the source of research evidence that public care agency leaders obtain. The process scale assesses how public care agency leaders evaluate the research evidence obtained and includes 3 subscales of self-assessment for validity and reliability of research evidence (10 items), reliance on others (5 items), and self-assessment for relevance (5 items). The output scale (20 items) assesses whether public care agency leaders eventually use the research evidence or ignore the evidence. The measurement responses use a 5-point Likert-type scale ranging from 1 (not at all) to 5 (all the time) for the items contained in the input scale and a 5-point Likert-type scale ranging from 1 (not important) to 5 (very important) for the items contained in the process and output scales. Each subscale measure and the total SIEU score are represented as average scores. Higher scores indicate higher agreement with the sources of evidence obtained for the input scale, more frequent evaluation of research evidence for the process scale, and greater use of research evidence for the output scale. SIEU has shown high internal consistency reliability (Cronbach α=.88) [41].

The PAP experience survey was developed for this study [11]. The questionnaire included 41 questions that were based on the potential PAP context, mechanism, and outcome configuration developed for the study [11,39]. These questions included both a Likert-type scale and open-ended questions. The survey items focused on the following four areas: (1) partnership purpose formulation (structure, goals, primary function, and agenda-setting process), (2) perceptions of partnership coalition building (mutual benefit, trust, convener’s role, leadership representation, role clarity, and conflict management), (3) perception of the PAP life cycle stage, and (4) public care agency leaders’ use of research evidence. We built and administered the web-based survey in the Research Electronic Data Capture [42], a secure web-based data collection tool that includes data entry forms and web surveying features.

A total of 48 public care agency leaders participated in the web-based SIEU survey scale [41], and 40 academic researchers and 26 public care agency leaders participated in the PAP experience survey. The survey response rates were 72% (48/67) for the SIEU survey, 48% (40/83) for academic researchers’ PAP experience survey, and 39% (26/67) for public care agency leaders’ PAP experience survey.

Analysis
The reliability of the SIEU was calculated using Cronbach α internal consistency for each of the subscales and the overall scale. Frequencies, percentages, and mean scores were calculated to identify (1) public care agency leaders’ and academic researchers’ ratings of alignment between PAP structure, goals, primary function, and agenda-setting process and their organizational structure, goals, primary function, and agenda-setting process by PAP life cycle stage (formed, matured, sustained, declining, and terminated); (2) public care agency leaders’ and academic researchers’ ratings of PAP coalition building (mutual benefits, trust, convener’s role, leadership representation, role clarity, and conflict management) by PAP life cycle stage; (3) public care agency leaders’ and academic researchers’ ratings of their partnership outcomes (identifying another issue to focus on and reformulate PAP purpose); and (4) public care agency leaders’ use of research evidence by the ratings of PAP life cycle stage, purpose formulation, and coalition building.

The original study design [11] was to recruit academic researchers and public care agency leaders in pairs. However, because of the low response rate for the PAP experience survey, we conducted group-level analysis for public care agency leaders and academic researchers, respectively, instead of conducting the analysis in pairs.
Results

Demographic Characteristics and Work Experience of Study Participants

As shown in Multimedia Appendix 1, public care agency leaders’ age and years of experience in the fields were distributed evenly. Of the public care agency leaders who answered the demographics and work experience questions, most (20/31, 65%) held a master’s degree. More than two-thirds of the public care agency leaders were women (21/31, 68%) and White (22/31, 71%). More than two-thirds of public care agency leaders (21/31, 68%) reported being at their current organizations for more than 10 years. More than three-fourths of public care agency leaders (24/31, 77%) had been involved in their current PAP for fewer than 10 years. The PAP roles they played were diverse and distributed evenly, and most (23/31, 74%) of public care agency leaders reported having been engaged in 5 or fewer PAPs.

As shown in Multimedia Appendix 2, academic researchers’ age and years of experience were also evenly distributed as were years at the current organization. Of the academic researchers who answered the demographic and work experience questions, most were women (30/40, 75%) and White (33/40, 83%) and held a doctoral degree (26/40, 65%). More than one-fourth of the academic researchers (11/40, 28%) had been involved with their current PAP for more than 10 years. The PAP roles they played were diverse, and only under one-third (11/40, 28%) identified their role as principal investigator, lead evaluator, and university lead.

PAP Life Cycle Stages and SIEU Scale Scores

The average total SIEU score was 3.1 (SD 0.81; range 0.9-4.1). The internal consistency reliability of the SIEU based on the study sample was high (Cronbach $\alpha=.89$). The mean score for the SIEU Input scale, the assessment of the source of research evidence that public care agency leaders obtain, was 2.9 (SD 0.46; range 1.8-3.9). The internal consistency reliability of the input scale was a Cronbach $\alpha$ value of .80. The mean SIEU process scale, the assessment of how public care agency leaders evaluate research evidence obtained, was 3.8 (SD 0.68; range 0.4-4.8). The internal consistency reliability of the process scale was a Cronbach $\alpha$ value of .85. The mean SIEU output scale, the assessment of public care agency leaders’ use of research evidence, was 3.1 (SD 0.74; range 0.2-3.9). The internal consistency reliability of the output scale was a Cronbach $\alpha$ value of .74.

As shown in Figure 1, 56% (15/26) of the public care agency leaders answered that their PAP was in a sustained stage, 22% (6/26) answered that their PAP was matured but did not reach a sustained stage yet, 11% (3/26) answered that their PAP was terminated at the time of the survey, 7% (2/26) answered that their PAP was declining, and 4% (1/26) answered that they were unsure about the stage of their PAP life cycle. None of the public care agency leaders answered that their PAP was formed but not matured yet.

For academic researchers, 45% (18/40) of the academic researchers answered that their PAP was in a sustained stage, 18% (7/40) answered that their PAP was matured but did not reach a sustained stage yet, 18% (7/40) answered that their PAP was formed but not reached a matured stage yet, 10% (4/40) answered that their PAP was declining, and another 10% (4/40) answered that their PAP was terminated.

For public care agency leaders’ use of research evidence (Figure 1), the public care agency leaders who answered that their PAP was declining had the highest SIEU output score (mean score 3.6, SD 0.42), followed by those who answered that their PAP was terminated (mean score 3.4, SD 0.13), those who answered that their PAP was mature but did not reach a sustained stage yet (mean score 3.3, SD 0.24), and those who answered that their PAP was sustained (mean score 3.3, SD 0.32).
Figure 1. Public care agency leaders’ perception of public–academic partnership life cycle stage and the Structured Interview for Evidence Use score. PAP: public–academic partnership; SIEU: Structured Interview for Evidence Use.

Perceptions of Purpose Formulation Context and Mechanism (Primary Function, Goals, Structure, and Agenda-Setting Process) and PAP Life Cycle Stage

As shown in Table 1, for PAPs in matured and sustained stages, only one public care agency leader in each group (1/6, 17% and 1/7, 7% of the PAPs, respectively) perceived the primary function of their PAP as perfectly aligned with the primary function of their organization. None of the public care agency leaders in PAPs declining and PAPs terminated perceived perfect alignment. A total of 4 academic researchers in formed, matured, and sustained PAPs (1/7, 14%; 1/7, 14%; and 2/17, 12% of the PAPs, respectively) perceived the primary function of partnership as perfectly aligned with the primary function of their organization.

Regarding the alignment of structures between PAP and partnering organizations, more than 86% (57/66) of both public care agency leaders and academic researchers answered that the structures were fairly well to perfectly aligned across all PAP life cycle stages. A total of 4 academic researchers in formed, matured, and sustained PAPs (1/7, 14%; 1/7, 14%; and 2/17, 12% of the PAPs, respectively) perceived the primary function of partnership as perfectly aligned with the primary function of their organization.

As shown in Table 2, 3 of the 5 public care agency leaders in the PAPs declining and terminated, perceived their PAP agenda-setting process as not at all or very little driven by the public care agency leaders. Academic researchers’ perception was similar; 3 of the 15 academic researchers in the PAPs formed, declining, and terminated perceived their PAP agenda-setting process as not at all or very little driven by public care agency leaders. More than 97% (30/31) of academic researchers in formed, matured, and sustained PAPs perceived their PAP agenda-setting process as driven by public care agency leaders. Almost half of the public care agency leaders (n=12) perceived very little of their PAP agenda process as driven by academic researchers, and this was consistent regardless of their perception of the PAP life cycle stage. The academic researchers’ perceptions were similar. Regardless of the PAP life cycle stage, almost half of academic researchers (n=18) perceived their PAP agenda-setting as not at all or very little driven by the researcher.
### Table 1. Public–academic partnership purpose formulation context: perception of alignment in primary function, structure, and organizational goals (public care agency leaders [N=26] and academic researchers [N=40]).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Formed, n (%)</th>
<th>Matured, n (%)</th>
<th>Sustained, n (%)</th>
<th>Declining, n (%)</th>
<th>Terminated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public care agency leaders (n=7)</td>
<td>Academic researchers (n=17)</td>
<td>Public care agency leaders (n=6)</td>
<td>Academic researchers (n=15)</td>
<td>Public care agency leaders (n=2)</td>
</tr>
<tr>
<td>Primary function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>N/A c</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very little</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Fairly well</td>
<td>N/A</td>
<td>2 (29)</td>
<td>2 (33)</td>
<td>0 (0)</td>
<td>3 (20)</td>
</tr>
<tr>
<td>Quite well</td>
<td>N/A</td>
<td>3 (43)</td>
<td>3 (42)</td>
<td>2 (13)</td>
<td>4 (24)</td>
</tr>
<tr>
<td>Very well</td>
<td>N/A</td>
<td>1 (14)</td>
<td>3 (50)</td>
<td>1 (14)</td>
<td>1 (67)</td>
</tr>
<tr>
<td>Perfectly</td>
<td>N/A</td>
<td>1 (14)</td>
<td>1 (17)</td>
<td>1 (14)</td>
<td>1 (67)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (67)</td>
</tr>
<tr>
<td>Structure alignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>N/A</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very little</td>
<td>N/A</td>
<td>1 (14)</td>
<td>2 (33)</td>
<td>1 (14)</td>
<td>3 (13)</td>
</tr>
<tr>
<td>Fairly well</td>
<td>N/A</td>
<td>1 (14)</td>
<td>2 (33)</td>
<td>1 (14)</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Quite well</td>
<td>N/A</td>
<td>3 (43)</td>
<td>3 (42)</td>
<td>3 (13)</td>
<td>3 (18)</td>
</tr>
<tr>
<td>Very well</td>
<td>N/A</td>
<td>2 (29)</td>
<td>5 (33)</td>
<td>3 (33)</td>
<td>5 (47)</td>
</tr>
<tr>
<td>Perfectly</td>
<td>N/A</td>
<td>2 (29)</td>
<td>5 (50)</td>
<td>1 (14)</td>
<td>1 (67)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (67)</td>
</tr>
<tr>
<td>Organizational goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very little</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Fairly well</td>
<td>N/A</td>
<td>2 (29)</td>
<td>4 (27)</td>
<td>1 (17)</td>
<td>4 (27)</td>
</tr>
<tr>
<td>Quite well</td>
<td>N/A</td>
<td>1 (14)</td>
<td>3 (43)</td>
<td>1 (17)</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Very well</td>
<td>N/A</td>
<td>3 (43)</td>
<td>4 (67)</td>
<td>3 (43)</td>
<td>3 (13)</td>
</tr>
<tr>
<td>Perfectly</td>
<td>N/A</td>
<td>1 (14)</td>
<td>1 (14)</td>
<td>1 (14)</td>
<td>1 (67)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

a For each cell, the within-column percentages of public care agency leaders’ and academic researchers’ perceptions are presented.
b Response missing for public care agency leaders (n=1); response missing for academic researchers (n=1).
c N/A: not applicable.
d Response missing for public care agency leaders (n=0); response missing for academic researchers (n=1).
e Response missing for public care agency leaders (n=0); response missing for academic researchers (n=1).
Table 2. Public–academic partnership (PAP) purpose formulation mechanism (agenda-setting process; public care agency leaders [N=26] and academic researchers [N=40])[^a].

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Formed, n (%)</th>
<th>Maturred, n (%)</th>
<th>Sustained, n (%)</th>
<th>Declining, n (%)</th>
<th>Terminated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public care agency leaders</strong></td>
<td><strong>Academic researchers</strong></td>
<td><strong>Academic researchers</strong></td>
<td><strong>Public care agency leaders</strong></td>
<td><strong>Academic researchers</strong></td>
<td><strong>Public care agency leaders</strong></td>
</tr>
<tr>
<td>(n=7)</td>
<td>(n=6)</td>
<td>(n=17)</td>
<td>(n=2)</td>
<td>(n=4)</td>
<td>(n=3)</td>
</tr>
<tr>
<td>Perceptions of PAP agenda driven by researchers[^b]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>N/A[^c]</td>
<td>1 (17)</td>
<td>1 (14)</td>
<td>3 (20)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very little</td>
<td>N/A</td>
<td>3 (43)</td>
<td>1 (14)</td>
<td>4 (27)</td>
<td>1 (50)</td>
</tr>
<tr>
<td>Fairly well</td>
<td>N/A</td>
<td>2 (29)</td>
<td>0 (0)</td>
<td>2 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Quite well</td>
<td>N/A</td>
<td>2 (29)</td>
<td>2 (29)</td>
<td>1 (7)</td>
<td>3 (24)</td>
</tr>
<tr>
<td>Very well</td>
<td>N/A</td>
<td>0 (0)</td>
<td>3 (43)</td>
<td>3 (20)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Perfectly</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

[^a]: For each cell, within-column percentages of public care agency leaders’ and academic researchers’ perception are presented, respectively.

[^b]: Response missing for public care agency leaders, (n=2); response missing for academic researchers (n=1).

[^c]: N/A: not applicable.

[^d]: Response missing for public care agency leaders, (n=0); response missing for academic researchers (n=1).

PAP Coalition Building Context (Convener’s Role, Leadership Representation, Role Clarity, and Conflict Management) and PAP Life Cycle Stage

As shown in [Table 3](#), most public care agency leaders and academic researchers in PAPs formed, matured, and sustained had a convener who gathered people together to carry out partnership processes, such as issue crystallization, partnership coalition building, and agenda-setting. In total, 3 of the 8 academic researchers in PAPs declining and terminated (2/4, 50% and 1/4, 25% of the PAPs, respectively) and 1 public care agency leader (1/3, 33%) in PAPs terminated answered that their PAPs were missing a convener.

Public care agency leaders’ perceptions of clear leadership representation and role clarity did not differ according to the PAP life cycle stage. Approximately 27% (4/15) of public care agency leaders in PAPs sustained answered that their PAP rarely or only occasionally had leadership representation and role clarity. Overall, 30 academic researchers (30/40, 75% of all academic researchers) answered that their PAP always had leadership representation and clear roles.

Most public care agency leaders and academic researchers answered that they experienced partnership conflict across all PAP life cycle stages, except for PAPs in a formed stage. Most public care agency leaders (up to 23/26, 88%; range 67%-100% across all PAP life cycle stages) answered that PAP members knew how to manage partnership conflicts. Most academic researchers in PAPs formed, matured, and sustained (up to 29/31, 94%; range 86%-100%) answered that their PAP members knew how to handle partnership conflicts. In total, 3 of the 8 academic researchers in PAPs declining and terminated answered that their PAP members knew how to handle partnership conflicts.

As shown in [Table 4](#), public care agency leaders’ trust in researchers, academic researchers’ trust in public care agency leaders, and perception of pursuing mutual benefit in partnership agenda-setting did not show meaningful patterns by the PAP life cycle stage. Most public care agency leaders (up to 23/26, 88%; range 67%-100% across all PAP life cycle stages) answered that PAP members knew how to manage partnership conflicts. Most academic researchers in PAPs formed, matured, and sustained (up to 29/31, 94%; range 86%-100%) answered that their PAP members knew how to handle partnership conflicts. In total, 3 of the 8 academic researchers in PAPs declining and terminated answered that their PAP members knew how to handle partnership conflicts.
mutual benefit in setting partnership agenda. PAPs sustained had the highest percentage of academic researchers (6/17, 35%) answering their PAP always pursued mutual benefits.

As shown in Table 5, across all PAP life cycle stages, most public care agency leaders and researchers answered that their PAP resulted in focusing on another issue. Academic researchers’ perception of their partnership leading to focus on another issue was the highest among PAPs matured (6/7, 86%), followed by PAPs sustained (14/17, 82%), PAPs declined (3/4, 75%), PAP terminated (2/4, 50%), and PAPs formed (3/7, 43%). More than two-thirds of the researchers (4/5, 67%) in PAPs matured answered that focusing on a new issue led to reformulating the PAP agenda-setting process. The majority of public care agency leaders (10/16, 63%) answered that the new issue did not result in reformulating the PAP agenda-setting process.

Table 3. Public–academic partnership (PAP) coalition building context (convener’s role, leadership representation, role clarity, and conflict management; public care agency leaders [N=26] and academic researchers [N=40]).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Formed, n (%)</th>
<th>Matured, n (%)</th>
<th>Sustained, n (%)</th>
<th>Declining, n (%)</th>
<th>Terminated, n (%)</th>
<th>Parameters</th>
<th>Formed, n (%)</th>
<th>Matured, n (%)</th>
<th>Sustained, n (%)</th>
<th>Declining, n (%)</th>
<th>Terminated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public care agency leaders (n=0)</td>
<td>Academic researchers (n=7)</td>
<td>Academic researchers (n=6)</td>
<td>Academic researchers (n=7)</td>
<td>Academic researchers (n=15)</td>
<td>Academic researchers (n=2)</td>
<td>Academic researchers (n=4)</td>
<td>Public care agency leaders (n=3)</td>
<td>Academic researchers (n=4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>N/A</td>
<td>6 (86)</td>
<td>5 (83)</td>
<td>5 (83)</td>
<td>9 (60)</td>
<td>14 (82)</td>
<td>2 (100)</td>
<td>2 (50)</td>
<td>2 (67)</td>
<td>1 (25)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>3 (20)</td>
<td>3 (18)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td></td>
</tr>
<tr>
<td>Used to have</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>2 (13)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td></td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Perception of PAP having a convener who plays the role of gathering people together&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Rarely</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (13)</td>
<td>1 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
</tr>
<tr>
<td>Occasionally</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (13)</td>
<td>2 (12)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td></td>
</tr>
<tr>
<td>Frequently</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>1 (14)</td>
<td>1 (7)</td>
<td>1 (6)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Very frequently</td>
<td>N/A</td>
<td>0 (0)</td>
<td>2 (33)</td>
<td>1 (14)</td>
<td>1 (7)</td>
<td>4 (26)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>N/A</td>
<td>7 (100)</td>
<td>3 (50)</td>
<td>5 (71)</td>
<td>8 (53)</td>
<td>9 (53)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Used to have</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Experience of PAP conflict&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Yes</td>
<td>N/A</td>
<td>1 (14)</td>
<td>3 (50)</td>
<td>4 (57)</td>
<td>10 (67)</td>
<td>8 (47)</td>
<td>1 (50)</td>
<td>4 (100)</td>
<td>2 (67)</td>
<td>4 (100)</td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>6 (86)</td>
<td>2 (33)</td>
<td>3 (43)</td>
<td>4 (27)</td>
<td>9 (53)</td>
<td>1 (50)</td>
<td>0 (0)</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Perception of PAP members knowing how to handle PAP conflicts&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Yes</td>
<td>N/A</td>
<td>7 (100)</td>
<td>4 (67)</td>
<td>6 (86)</td>
<td>14 (93)</td>
<td>16 (94)</td>
<td>2 (100)</td>
<td>1 (25)</td>
<td>3 (100)</td>
<td>2 (50)</td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>0 (0)</td>
<td>2 (33)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (6)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td></td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (14)</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>For each cell, the within-column percentages of public care agency leaders’ and academic researchers’ perceptions are presented.
<sup>b</sup>Response missing for public care agency leaders (n=0); response missing for academic researchers (n=2).
<sup>c</sup>N/A: not applicable.
<sup>d</sup>Response missing for public care agency leaders (n=0); response missing for academic researchers (n=1).
<sup>e</sup>Response missing for public care agency leaders (n=0); response missing for academic researchers (n=1).
<sup>f</sup>Response missing for public care agency leaders (n=0); response missing for academic researchers (n=1).
Table 4. Public–academic partnership (PAP) coalition building mechanism (mutual benefit and trust in PAP agenda-setting; public care agency leaders [N=26] and academic researchers [N=40])

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Formed, n (%)</th>
<th>Matured, n (%)</th>
<th>Sustained, n (%)</th>
<th>Declining, n (%)</th>
<th>Terminated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public care agency leaders (n=7)</td>
<td>Academic researchers (n=7)</td>
<td>Public care agency leaders (n=15)</td>
<td>Academic researchers (n=17)</td>
<td>Public care agency leaders (n=2)</td>
</tr>
<tr>
<td>Perceived mutual benefit in PAP agenda setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rarely</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Occasionally</td>
<td>N/A</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Frequently</td>
<td>N/A</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>3 (50)</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Very frequently</td>
<td>N/A</td>
<td>2 (29)</td>
<td>0 (0)</td>
<td>2 (33)</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Always</td>
<td>N/A</td>
<td>3 (43)</td>
<td>4 (67)</td>
<td>1 (17)</td>
<td>7 (47)</td>
</tr>
<tr>
<td>Used to pursue mutual benefit</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Perceived level of trust academic researchers have for public care agency leaders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>5 (86)</td>
<td>4 (67)</td>
<td>5 (71)</td>
<td>7 (47)</td>
</tr>
<tr>
<td>Moderate</td>
<td>N/A</td>
<td>1 (14)</td>
<td>1 (17)</td>
<td>1 (14)</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Low</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Used to have high level of trust</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>1 (14)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Perceived level of trust public care agency leaders have for academic researchers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>5 (71)</td>
<td>5 (83)</td>
<td>5 (63)</td>
<td>7 (47)</td>
</tr>
<tr>
<td>Moderate</td>
<td>N/A</td>
<td>2 (29)</td>
<td>0 (0)</td>
<td>1 (13)</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Low</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Used to have high level of trust</td>
<td>N/A</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (13)</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

a For each cell, the within-column percentages of public care agency leaders’ and academic researchers’ perceptions are presented.
b Response missing for public care agency leaders (n=0); response missing for academic researchers (n=2).
c N/A: not applicable.
d Response missing for public care agency leaders (n=0); response missing for academic researchers (n=2).
e Response missing for public care agency leaders (n=0); response missing for academic researchers (n=0).
Table 5. Public–academic partnership (PAP) purpose formulation and coalition building outcome (new issue to focus and reformulation of PAP agenda-setting process; public care agency leaders [N=26] and academic researchers [N=40]).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Formed, n (%)</th>
<th>Matured, n (%)</th>
<th>Sustained, n (%)</th>
<th>Declining, n (%)</th>
<th>Terminated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public care agency leaders (n=7)</td>
<td>Academic researchers (n=4)</td>
<td>Public care agency leaders (n=6)</td>
<td>Academic researchers (n=17)</td>
<td>Public care agency leaders (n=4)</td>
</tr>
<tr>
<td>Terminated</td>
<td>2 (8)</td>
<td>6 (75)</td>
<td>1 (50)</td>
<td>1 (75)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Declining</td>
<td>3 (50)</td>
<td>3 (50)</td>
<td>6 (86)</td>
<td>14 (82)</td>
<td>1 (50)</td>
</tr>
<tr>
<td>Sustained</td>
<td>10 (67)</td>
<td>4 (27)</td>
<td>2 (12)</td>
<td>1 (50)</td>
<td>1 (25)</td>
</tr>
<tr>
<td>Formed</td>
<td>2 (50)</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (50)</td>
<td>2 (64)</td>
</tr>
<tr>
<td>Academic researchers (n=4)</td>
<td>Academic researchers (n=4)</td>
<td>Academic researchers (n=3)</td>
<td>Academic researchers (n=4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception of PAP leading to focus on another issue(^b)</td>
<td>Yes</td>
<td>No</td>
<td>Do not know or unsure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>N/A(^c)</td>
<td>3 (43)</td>
<td>3 (50)</td>
<td>6 (86)</td>
<td>10 (67)</td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>4 (57)</td>
<td>2 (33)</td>
<td>1 (14)</td>
<td>4 (27)</td>
</tr>
<tr>
<td>Do not know or unsure</td>
<td>N/A</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
</tbody>
</table>

\(^a\)For each cell, the within-column percentages of public care agency leaders’ and academic researchers’ perceptions are presented.

\(^b\)Response missing for public care agency leaders (n=0); response missing for academic researchers (n=1).

\(^c\)N/A: not applicable.

\(^d\)Response missing for public care agency leaders (n=10); response missing for academic researchers (n=12).

PAP Purpose Formulation and Coalition Building and Public Care Agency Leaders’ Use of Research Evidence

Figures 1-9 present public care agency leaders’ perceptions of PAP purpose formulation and coalition building and their use of research evidence. The average SIEU output scale score that indicates public care agency leaders’ actual use of research evidence was the highest among the PAPs declining followed by PAPs terminated, PAPs formed, and PAPs matured (3.6, SD 0.42; 3.4, SD 0.13; 3.3, SD 0.32; and 3.3, SD 0.24, respectively). The average SIEU output scale score was higher in PAPs, which resulted in another issue to focus compared with the score of PAPs without issue recrystallization (SIEU scores 3.4, SD 0.28 vs 3.1, SD 0.25).

On the other hand, the SIEU output scale score did not show a correlated pattern with public care agency leaders’ perceptions of the agenda-setting process. Public care agency leaders who reported that their partnering researchers used to have trust in PAP leaders (public care agency leaders) showed the highest average SIEU output scale score (3.5, SD 0.21). The SIEU output scale scores did not show correlated pattern with the public care agency leaders’ perception of PAP seeking mutual benefit.
Figure 2. Public care agency leaders’ perception of goal alignment and the Structured Interview for Evidence Use score. SIEU: Structured Interview for Evidence Use.

Figure 3. Public care agency leaders’ perception of primary function alignment and the Structured Interview for Evidence Use score. SIEU: Structured Interview for Evidence Use.
Figure 4. Public care agency leaders’ perception of structure alignment and the Structured Interview for Evidence Use score. SIEU: Structured Interview for Evidence Use.

Figure 5. Public care agency leaders’ perception of agenda-setting driven by public care agency leaders and the Structured Interview for Evidence Use score. SIEU: Structured Interview for Evidence Use.
**Figure 6.** Public care agency leaders’ perception of agenda-setting driven by researchers and the Structured Interview for Evidence Use score. SIEU: Structured Interview for Evidence Use.

**Figure 7.** Public care agency leaders’ response for partnership issue crystallization and the Structured Interview for Evidence Use score. PAP: public–academic partnership; SIEU: Structured Interview for Evidence Use.
Figure 8. Public care agency leaders’ response for partnership pursuing mutual benefits and the Structured Interview for Evidence Use score. PAP: public–academic partnership; SIEU: Structured Interview for Evidence Use.

Figure 9. Public care agency leaders’ perception of level of trust researchers have for the public care agency leaders and the Structured Interview for Evidence Use score. PAP: public–academic partnership; SIEU: Structured Interview for Evidence Use.
**Discussion**

**Purpose Formulation, Coalition Building, and PAP Life Cycle Stages**

The study findings show that overall, PAP purpose formulation including goals, primary function and structure, and partnership coalition building, including mutual benefits, trust, convener’s role, leadership representation, role clarity, and conflict management, are important contexts and mechanisms for PAPs to evolve through life cycle stages. For the partnership contexts and mechanisms, PAPs matured were perceived more positively than PAPs formed, and PAPs sustained were perceived more positively than PAPs matured by public care agency leaders and academic researchers. However, not all the contexts and mechanisms of purpose formulation and coalition building showed evolving through the PAP life cycle stages.

Most public care agency leaders and academic researchers in PAPs formed, matured, and sustained perceived the context of partnership purpose formulation as well aligned with those of their organization. Public care agency leaders and academic researchers in PAPs declining and terminated perceived a low level of alignment in the context. This echoes the findings from studies focused on PAPs in other fields, such as public health insurance [31], environmental health [32], health care delivery [30], child welfare and mental health services [10], and general and adult mental health care [2,4-6,37] in which successful PAPs were reported to have aligned structure, goals, and agenda-setting process.

More than 40% (30/66) of public care agency leaders and academic researchers in PAPs sustained perceived that PAP agenda-setting was not at all or little driven by them. Particularly, more than one-third of researchers perceived that their PAP agenda-setting was not at all driven by academic researchers. More than 50% (25/45) of public care agency leaders and academic researchers in PAP matured and sustained perceived their PAP as always having leadership presentation and role clarity. As demonstrated in previous studies [10,34], a continuous role clarity process that responds to changing environments and needs of the mental health, child welfare, and public health fields is important for PAPs to sustain. PAPs sustained are likely to have overcome periodic leadership shifts and changes in the political environment, successfully engaging new leaders in the partnership process and continuously clarifying the roles of the members of PAP [10].

Effective conflict management skills have been shown to be important in building successful PAPs in health care delivery [30]. Most public care agency leaders and researchers experienced partnership conflict regardless of the PAP life cycle stage, except for the researchers in PAPs formed. Most public care agency leaders and academic researchers in PAPs formed, matured, and sustained reported that their PAP members knew how to handle partnership conflicts.

We found that most PAPs across all life cycle stages crystallized another issue, but the issue of crystallization did not lead to purpose reformulation for most PAPs. Although partnerships are expected to constantly review and reformulate purpose and scan their environmental changes to increase their sustainability [21], it is possible that focusing on another issue does not require changes in the PAP agenda-setting process. We did not have information on whether the new issue crystallization required PAPs to reformulate partnership purpose. Further research on specific PAP context and mechanisms that result in PAP purpose reformulation will lead to gaining an in-depth understanding.

**Public Care Agency Leaders’ Use of Research Evidence by Perception of PAP Purpose Formation, Coalition Building, and PAP Life Cycle Stage**

Supporting the previous research [29] on context and mechanisms for successful PAPs, our study found that developing trusting relationships with public care agency leaders and continuously crystallizing PAP issues play an important role in not only increasing PAP sustainability but also fostering public care agency leaders’ use of research evidence. Public care agency leaders using research evidence may be more open to new ideas proposed by academic researchers and actively pursue issue crystallization. PAPs that continuously crystallize issues are also likely to lead public care agency leaders to be frequently exposed to research evidence. Public care agency leaders who reported their PAP as having a high level of trust in their partnering researchers also showed greater use of research evidence.

Unlike the previous research in health care delivery [30] that reports identifying clear and aligned goals as promoting partners’ prioritization of their work and eventual use of evidence, our study did not find greater use of research evidence among public care agency leaders who perceived their PAP goals, primary function, and structure well aligned with their organizational goals, primary function, and structure. Previous research on health care delivery [30] and public health [36] have reported a positive relationship between PAPs seeking mutual benefit and public care agency leaders’ use of research evidence. However, our study did not find the positive relationship. Public care agency leaders’ use of research evidence did not show a consistent pattern by the PAP life cycle stage. Public care agency leaders who perceived their PAP as declining showed the highest level of use of research evidence. This may be attributed to the small sample size, and further research is warranted. Future research with a larger study sample and mixed methods will provide further insights.

**Limitations**

Our study has a few limitations. The number of public care agency leaders who participated in the PAP experience survey was limited to 26. We described in the informed consent that information provided by study participants would remain in a secure web-based database that only the key research staff could access, and that data would be analyzed at the aggregate level. Despite the statement of confidentiality and privacy protection written in the informed consent, the response rate from public care agency leaders was low. Some of the contexts and mechanisms of PAP purpose formulation and coalition building not varying by PAP life cycle stage may be attributed to the small sample size. Academic researchers’ and public care agency leaders’ PAP partnership experience were not analyzed in pairs because of the small sample size. Thus, our findings do not.
reflect the concordance level in the perception of academic researchers and public care agency leaders in pairs. The study findings may reflect social desirability bias from the respondents. For example, as noted by Ross et al [36], researchers may have reported on positive aspects of the relationships with public care agency leaders to avoid damaging connections, and policy makers might have reported stronger reliance on evidence use because of public emphasis on evidence use. Some of the PAP contexts, such as funding opportunities and mental health and child welfare policies at the federal and state levels, are expected to influence PAP sustainability and public care agency leaders’ use of research evidence. In this study, we focused on the contexts and mechanisms that can be applied to all PAPs in the fields instead of reviewing and interpreting PAP-specific contexts. A case analysis that incorporates PAP-specific contexts along with the purpose formulation and partnership coalition building can provide in-depth insights.

Conclusions
Understanding factors that promote successful PAPs and evidence use by policy makers has the potential to improve outcomes for vulnerable youth populations served by public mental health and child welfare systems in the United States. PAPs declining can revive through making changes to adapt to continuously changing environment. Our study findings suggest that continuous trust cultivation through ongoing and clear communication and continuous issue crystallization may promote public care agency leaders’ use of research evidence. Academic researchers’ efforts to build trust with public care agency leaders and constantly formulate issues to meet the needs of public care agency leaders who constantly experience changes in the public care environment are essential. To promote mutual benefits that link to the use of research evidence, public care agencies should establish clear research and evaluation guidelines to inform researchers of expectations when initiating and forming PAPs.

Few studies have examined PAPs in the mental health and child welfare fields despite the frequent use of PAPs. Recently, there has been increased attention to PAPs in other related fields such as health care, with rapid advancement of science such as health information technology [43]. PAPs play an important role in translating research findings into innovative policies and practices. We urge academic researchers and public care agency leaders in the fields of mental health and child welfare to pay greater attention to further understanding the partnership context and mechanisms that promote innovative evidence-based policy and practice.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Public care agency leaders’ demographics and work experience.
[DOCX File , 16 KB - humanfactors_v9i1e29288_app1.docx ]

Multimedia Appendix 2
Academic researchers’ demographics and work experience.
[DOCX File , 16 KB - humanfactors_v9i1e29288_app2.docx ]

References


14. Findings from the California Youth Transitions to Adulthood Study (CaYOUTH): conditions of youth at age 19. Chapin Hall at the University of Chicago. 2016. URL: https://www.chapinhall.org/wp-content/uploads/CY_YT_RE0516_4-1.pdf [accessed 2021-12-17]


17. How the child welfare system works. Children's Bureau. URL: https://www.childwelfare.gov/pubs/CPSPDFs/cpswork.pdf [accessed 2021-12-10]


20. Funding and characteristics of single state agencies for substance abuse services and state mental health agencies, 2015.


40. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]


42. REDCap homepage. Research Electric Data Capture. URL: https://www.project-redcap.org/ [accessed 2021-12-17]

An Interactive Voice Response System to Increase Physical Activity and Prevent Cancer in the Rural Alabama Black Belt: Design and Usability Study

Mohanraj Thirumalai, MS, MEng, PhD; Nashira Brown, MS; Soumya Niranjan, BPharm, MS, PhD; Sh’Nese Townsend, MHA; Mary Anne Powell, MA; Erica Schleicher; Venkatadri Raparla, MS; Robert Oster, PhD; Wendy Demark-Wahnefried, PhD; Dori Pekmezi, PhD

1Department of Health Services Administration, School of Health Professions, University of Birmingham at Alabama, Birmingham, AL, United States
2Department of Nutrition Sciences, School of Health Professions, University of Birmingham at Alabama, Birmingham, AL, United States
3Department of Health Behavior, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, United States
4University of Alabama at Birmingham/Lakeshore Research Collaborative, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States
5Division of Preventive Medicine, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Mohanraj Thirumalai, MS, MEng, PhD
Department of Health Services Administration
School of Health Professions
University of Birmingham at Alabama
SHPB #590E
1716 9th Avenue South
Birmingham, AL, 35244
United States
Phone: 1 2059347189
Email: mohanraj@uab.edu

Abstract

Background: Increased physical activity (PA) levels are associated with reduced risk and improved survival for several cancers; however, most Americans engage in less than the recommended levels of PA. Using interactive voice response (IVR) systems to provide personalized health education and counseling may represent a high-reach, low-cost strategy for addressing physical inactivity and cancer disparities in disproportionately burdened rural regions. However, there has been a paucity of research conducted in this area to date.

Objective: The aim of this study is to design, develop, and test the usability of an IVR system aimed at increasing PA levels in the rural Alabama Black Belt.

Methods: A pilot version of the IVR system was used to assess initial feasibility and acceptability. Detailed exit interviews were conducted to elicit participant feedback, which helped inform the development of a substantially upgraded in-house IVR system. This refined IVR system was then subjected to a sequential explanatory mixed methods evaluation. Participating rural county coordinators and research staff (N=10) tested the usability of the IVR system features for 2 weeks and then completed the System Usability Scale and qualitative semistructured interviews.

Results: The study sample comprised mostly African American people, women, rural county coordinators, and research staff (N=10). Participants rated the IVR system with a mean score of 81 (SD 5) on the System Usability Scale, implying excellent usability. In total, 5 overarching themes emerged from the qualitative interviews: likes or dislikes of the intervention, barriers to or facilitators of PA, technical difficulties, quality of calls, and suggestions for intervention improvement. Message framing on step feedback, call completion incentives, and incremental goal-setting challenges were areas identified for improvement. The positive areas highlighted in the interviews included the personalized call schedules, flexibility to call in or receive a call, ability to make up for missed calls, narration, and PA tips.
Conclusions: The usability testing and feedback received from the rural county coordinators and research staff helped inform a final round of refinement to the IVR system before use in a large randomized controlled trial. This study stresses the importance of usability testing of all digital health interventions and the benefits it can offer to the intervention.

(JMIR Hum Factors 2022;9(1):e29494) doi:10.2196/29494

KEYWORDS
interactive voice response systems; usability; exercise; physical activity; rural health; telehealth

Introduction

Background
Automated telephone-based intervention strategies may be key to overcoming the numerous barriers to physical activity (PA) promotion and cancer control in the Alabama Black Belt, a rural region named for its rich soil but whose population is at increased risk for sedentary lifestyles and related cancer disparities [1]. Low literacy, poverty, lack of transportation, cultural preferences, and distance from PA facilities often impede access to PA information and resources in this region [2]. Interactive voice response (IVR) systems allow users to interact by pressing keys on the telephone keypad and can be effective in targeting behavior change [3]. The recent National Health Interview Survey estimates that only 0.7% of the population in the United States is phoneless [4], thereby demonstrating the potential for a wider reach of IVR interventions. Moreover, IVRs do not require clinic visits, high literacy, or access to costly technology [5,6].

In response, we have developed an IVR-delivered PA intervention that is currently being tested in a large randomized controlled trial (RCT) in 6 rural Black Belt counties of Alabama. This paper describes the process that led us to the design of the IVR system and the results of the usability testing that was conducted before the commencement of the RCT.

As with any intervention, particularly digital health interventions, examination of the usability of the developed intervention before the actual deployment of the intervention is vital [7]. With IVR systems featuring only voice-based output and keypad-based input, a seamless user experience can indeed be tricky [8-10]. IVR-based intervention systems can pose more challenges than simple IVR data collection systems, as IVR-based intervention systems need to focus on achieving minimal information navigation time, while featuring maximal information relevance and capacity [8].

Objectives
Our proposed study aims to target rural Black Belt counties of Alabama that are marked by low literacy and education levels [1,2]. Although there is a body of work focused on the usability of IVR systems [8-10], there is limited research on the usability of IVR systems for rural settings and underserved populations. This limited body of literature has used surveys and interviews to evaluate the usability of IVR systems. This study seeks to fill this gap in the literature by reporting our development methodology, system features, and explanatory sequential mixed methods design to assess the usability of the IVR system. Our hypotheses are that most participants in this usability study will rate the usability of the IVR system favorably and provide useful suggestions for further improvements during interviews.

Methods

Parent Study Overview
The parent study (R01CA233550) is an ongoing RCT (N=240) comparing a Deep South IVR-Supported Active Lifestyle (DIAL) intervention with a waitlist control among underactive adults residing in 6 rural Alabama counties [11]. On the basis of the social cognitive theory (SCT) [12], this study extends an IVR-supported PA intervention that targets key SCT constructs (self-regulation, self-efficacy, enjoyment, outcome expectations, and social support) through IVR counseling calls. The participants are provided pedometers (Accusplit AX2790MV) and Fitbit activity monitors (model: Inspire) to record daily steps and receive progress feedback via the IVR PA-tracking and goal-setting calls.

The number of calls in a week tapers as participants progress through the intervention (from daily calls in months 0-3 to twice per week in months 4-6 and weekly in months 7-12), and the content of the calls vary based on specific days of the intervention.

Iterative IVR System Design

Piloting a Beta Version of IVR
A previous pilot study (R03CA177538) tested a beta version of this IVR system with a convenience sample (N=63) [13-15]. Findings from this trial supported the feasibility and acceptability of the approach and helped further refine the technology and theory-driven intervention components in preparation for extension to rural populations. More specifically, the findings yielded the need for IVR-initiated calls as opposed to only participant-initiated calls, specific targeting of unchanged SCT constructs and incorporating multi-level strategies (incremental goal-setting and county coordinator support) for increased support, accountability, and sustainability [14].

The IVR system used in this pilot study was a commercial IVR system and posed several limitations. First, all voice clips were prerecorded by voice narrators and uploaded. Second, the system only worked by participants calling into the system and did not offer a way for the system to initiate calls. The commercial system also posed limitations in terms of dynamic tailored questions that used earlier responses to frame newer questions as the call progressed.

Upgrading and Refining the IVR System
In response to this pilot study feedback, we developed a completely homegrown IVR system for the parent RCT using...
more up-to-date technology. This new system was hosted on a
Linux server, powered by an Apache web server, programmed
using the Laravel framework (a hypertext preprocessor–based
rapid development framework), data stored using a MySQL
database, and connected to Twilio for telephony.

Although the system was being developed, we conducted focus
groups with multiple stakeholder groups (rural county
coordinators and research staff from the University of Alabama
at Birmingham [UAB] O’Neal Comprehensive Cancer Center
Community Outreach and Engagement Office). For the focus
groups, we generated 3 sample voice clips of intervention
messages using Amazon Polly, a text-to-speech engine, and
presented the 3 sample voice clips to our stakeholders. Amazon
Polly is capable of close to human-like voices, which resulted
in stakeholders preferring Amazon Polly voices over prerecorded
human voices. This choice of Amazon Polly voices also allows
for the use of different tones and genders for the voices during the
calls and avoids the extensive time and financial costs
associated with rerecording message libraries with human
narrators every time an edit is made to the content.

The focus group participants also provided feedback on
incoming versus outgoing calls, the preferred procedure to
handle missed calls, and other support strategies. More
specifically, they felt that their community members and
potential participants would appreciate the flexibility and
convenience of bidirectional calls and the option to fill in missed
call data at later dates. The need to be able to change phone
numbers and allow incoming calls from new (unregistered)
numbers was stressed. Support strategies, such as brief
counseling sessions during in-person data collection and offering
Fitbit devices were also suggested.

The system development was conducted in an agile fashion,
with regular demonstrations to the rural country coordinators
from the UAB O’Neal Comprehensive Cancer Center
Community Outreach and Engagement Office. Their feedback
regarding the speed of the voice clips, pauses between sentences,
 pauses in sentences, length of the phone call, reading level of
the language used in the calls, and logical flow of the content
resulted in numerous edits. The system included a participant
call completion incentive mechanism that awarded the
participants a minimum of US $0.25 for each call completed.
However, the incentive amount became US $0.50 when the
participant completed 7 preceding calls, with the incentive
falling back to US $0.25 when a call was not completed.

The development phase concluded with the core project staff
(DP, MT, ST, and VR) pilot-testing the revised system to
identify and fix any problems. Some examples of the problems
identified and fixed include system expecting responses within
5 seconds, incorrect feedback messages, and outgoing calls not
being placed as scheduled. After this, a formal usability test
was conducted as detailed in the following section. Finally, the
system went through another round of iterative refinements
based on the findings from the usability testing. The details of
the resultant system are presented in the Results section.

Usability Testing

Study Design

This study incorporated an explanatory sequential mixed
methods design to assess the usability of an IVR phone
counseling system that will be extended to physically inactive
residents in 6 rural Alabama counties (Hale, Choctaw, Greene,
Marengo, Dallas, and Sumter). Demographics were assessed at
baseline. System usability and semistructured interviews were
conducted at the 2-week follow-up.

Participants

The sample for usability testing comprised 10 rural county
coordinators and research staff affiliated with the UAB O’Neal
Comprehensive Cancer Center Community Outreach and
Engagement Office who would later serve a critical role in
recruitment, assessment, and intervention delivery for the RCT
study but had yet to be exposed to the newly developed IVR
system.

Procedures

Each participant completed a one-on-one orientation via Zoom
with the DIAL program manager or principal investigator.
During the session, the participants were given an overview of
the usability study protocols and the IVR system, completed an
initial IVR call with the research team, and asked questions.

Following orientation, the participants began wearing a
study-provided pedometer or an approved personal activity
monitor (ie, Fitbit or Apple Watch) and receiving daily IVR
calls from DIAL for 2 weeks. The participants received all 3
types of IVR calls: PA-tracking, goal-setting, and counseling
calls. For tracking calls, the participants answered PA questions
reported pedometer use, steps per day, and any
moderate-intensity PA in the past 24 hours) and received PA
tips and feedback. Tracking calls lasted approximately 1 minute
per call. During the counseling calls, the participants answered
PA questions and additional questions covering PA self-efficacy,
enjoyment, outcome expectations, and social support. Moreover,
they received tailored feedback on these psychosocial variables
based on their individual responses to these questions.
Counseling calls lasted approximately 10 minutes per call.
Goal-setting calls allowed the participants to set their own step
goal or increase their current step goal by 500 steps for the
upcoming week. Goal-setting calls lasted approximately 5
minutes per call. In the 2-week period, the participants received
1 call per day, with a total of 1 goal-setting call, 1 counseling
 call, and 12 PA-tracking calls.

Quantitative Measures

Demographics

Participant demographics, including age, gender, educational
attainment, race and ethnicity, household income, employment,
marital status, and number of children living at home were
assessed at baseline.

Survey Items

At follow-up, the participants completed the System Usability
Survey (10 items) on the web via Qualtrics XM (Qualtrics)
combined with 4 more project-specific items. All 14 items were
aimed at assessing how the participants felt about the phone counseling system after using it for 2 weeks. The participants responded to the statement—Please select the answer that best expresses how you feel about each statement after using the phone counseling system over the past 2 weeks—for items such as I think I would like to use this phone counseling system frequently, I thought the phone counseling system was easy to use, I felt very confident using the phone counseling system, and I needed to learn a lot of things before I could get going with this phone counseling system. The 4 project-specific items were worded as How likely are you to recommend this system to others?, Did you receive your calls at the scheduled time?, What gender was the voice on your calls?, and Did you use the study-provided pedometer to track your steps?

**Quantitative Analysis**

All quantitative data collected during this study were descriptively analyzed. Microsoft Excel was used for all the quantitative analyses.

**Qualitative Methodology**

After 2 weeks of receiving calls and completing the quantitative survey, all 10 rural county coordinators and research staff participated in one-on-one, semistructured interviews conducted via Zoom regarding their experiences with the calls and how usability could be improved before implementing the IVR for the RCT. The semistructured interview guide was developed by coauthors (DP and SN) and included questions regarding motivation to exercise, likes and dislikes of the calls, specific call features that could motivate or demotivate individuals, technical aspects of the IVR call, and suggestions for improvement. To ensure consistency, all interviews were conducted in July 2020 by 1 member of the study team (SN) with expertise and experience in qualitative interviewing. SN is not involved in any aspect of the broader RCT or technology design and development and was engaged to serve as a neutral evaluator for the purpose of this usability evaluation.

**Results**

**Participant Characteristics**

Participant characteristics are shown in Table 1. The total sample included 10 participants with an average age of 48.7 (SD 18.6) years. The sample comprised largely female (8/10, 80%), Black (8/10, 80%), and non-Hispanic or Latino (9/10, 90%) participants, and had no children living at home (9/10, 90%). Most reported completing college (6/10, 60%) and either full-time or part-time employment at the time of usability testing. Half of the sample (5/10, 50%) reported <US $50,000 annual household income, and only 30% (3/10) of the participants reported never being married.
Table 1. Demographic characteristics of the usability testing participants (N=10).

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>48.7 (18.6)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8 (80)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>8 (80)</td>
</tr>
<tr>
<td>Non-Hispanic or Latino</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (10)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>4 (40)</td>
</tr>
<tr>
<td>College graduate</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Postgraduate work</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Household annual income (US $), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;50,000</td>
<td>5 (50)</td>
</tr>
<tr>
<td>≥50,000</td>
<td>5 (50)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Single (never married)</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Married</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Divorced</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Children living at home, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>9 (90)</td>
</tr>
<tr>
<td>≥1</td>
<td>1 (10)</td>
</tr>
</tbody>
</table>

Quantitative Results

The usability testing survey that was conducted after 2 weeks of IVR system use yielded positive results (Table 2). All participants (10/10, 100%) agreed that the IVR system was easy to use without the need for technical assistance or extensive learning, and most (7/10, 70%) would recommend the IVR system to others. The participants were confident in using IVR (8/10, 80%), and 70% (7/10) would like to use IVR frequently. Very few participants found the IVR system cumbersome (2/10, 20%) or confusing (3/10, 30%), and only 10% (1/10) of the participants found the IVR system to be unnecessarily complex. In terms of functionality, 70% (7/10) of the participants agreed that the various functions of the IVR system were well-integrated. The participants (7/10, 70%) reported receiving their calls at the scheduled time, and 90% (9/10) reported a female voice on their calls. Only 40% (4/10) of the participants reported wearing the study pedometer; however, of the 60% (6/10) who did not wear the study pedometer, 50% (3/6) used an Apple Watch and 50% (3/6) used a Fitbit Inspire. To numerically interpret the usability of the system, the standardized System Usability Scale scoring procedure was used [19]. This resulted in an average score of 81 (SD 5). Previous research indicates that a System Usability Scale score of >68 can be considered as above-average usability. This score of 81 translates to an excellent usability rating [20].
Table 2. Usability testing survey results (N=10).

<table>
<thead>
<tr>
<th>Statement and answers</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think I would like to use this phone counseling system frequently.</td>
<td></td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>3 (30)</td>
</tr>
<tr>
<td>I found the phone counseling system unnecessarily complex.</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>1 (10)</td>
</tr>
<tr>
<td>I thought the phone counseling system was easy to use.</td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td>5 (50)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>5 (50)</td>
</tr>
<tr>
<td>I think that I would need the support of a technical person to be able to use this system.</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>3 (30)</td>
</tr>
<tr>
<td>I found the various functions in this phone counseling system were well-integrated.</td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>3 (30)</td>
</tr>
<tr>
<td>I thought there was too much inconsistency in this phone counseling system.</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>1 (10)</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use this phone counseling system very quickly.</td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>6 (60)</td>
</tr>
<tr>
<td>I found the phone counseling system very cumbersome to use.</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>1 (10)</td>
</tr>
<tr>
<td>I felt very confident using the phone counseling system.</td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>1 (10)</td>
</tr>
<tr>
<td>I needed to learn a lot of things before I could get going with this phone counseling system.</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>5 (50)</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>5 (50)</td>
</tr>
<tr>
<td>How likely are you to recommend this system to others? (Scale of 0-10)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (20)</td>
</tr>
<tr>
<td>6</td>
<td>1 (10)</td>
</tr>
<tr>
<td>8</td>
<td>4 (40)</td>
</tr>
</tbody>
</table>
Statement and answers\textsuperscript{a} & Participants, n (%) \\
\hline
9 & 2 (20) \\
10 & 1 (10) \\
\hline

Did you receive your calls at the scheduled time? \\
Yes & 7 (70) \\
No & 3 (30) \\
\hline

What gender was the voice on your calls? \\
Female & 9 (90) \\
Both male and female & 1 (10) \\
\hline

Did you use the study-provided pedometer to track your steps? \\
Yes & 4 (40) \\
No & 6 (60) \\
Apple Watch & 3 (30) \\
Fitbit Inspire & 3 (30) \\
\hline

\textsuperscript{a}Please select the answer that best expresses how you feel about each statement after using the phone counseling system over the past 2 weeks.

Qualitative Results

A total of 5 overarching themes emerged: (1) likes or dislikes of the intervention, (2) barriers to or facilitators of PA, (3) technical difficulties, (4) quality of the calls, and (5) suggestions for improvement of the intervention.

Likes and Dislikes About the IVR Intervention

When asked what they liked about the DIAL intervention, several participants stated that the phone calls motivated them to exercise and kept them accountable:

\begin{quote}
I wasn’t as active, but after I went through the calls, I became more active and aware, and I was becoming used to the calls, and I was looking forward to the calls, and I was looking forward to the motivational tips.
\end{quote}

I think because it held me accountable. The accountability to hear what I had accomplished and what I not accomplished, that adds extra value because it almost puts a mirror in front of your face, and says, “Look.” Sometimes it’s very difficult to look at that mirror, and say, “This is what I have or have not done.”

The participants also appreciated the flexibility of the (new) bidirectional call format:

\begin{quote}
One thing I did like was that, for instance, if I did not make my call, I had the opportunity to call back. That was good.

You have the different options and different times of calling, I think that’s good for the people that’s busy. So if they miss the call, they can call back, or the system will call them back, but if they need to change their time of the call, then they able to do that.
\end{quote}

Finally, the participants looked forward to the PA tips at the end of the call:

\begin{quote}
I think motivating tips at the end, they were good. I knew them already, but I listened to them. So I think that they were good for people that’s just starting out with their health journey.
\end{quote}

Regarding dislikes, the participants expressed concerns that specific step feedback messages were negative and stern. For example, when <10,000 steps per day were reported in the PA-tracking call, the participants received the following feedback:

\begin{quote}
Thanks for reporting your steps. You did not meet the DIAL study step goal of 10,000 steps per day yet, but you are on your way. Keep making small increases until you get there.
\end{quote}

The participants had strong reactions to this feedback and compared it to a slap in the face:

\begin{quote}
It would be a little discouraging to hear that every day, “You didn’t meet your 10,000 steps goal. You did not meet the goal. You did not meet the goal.”
\end{quote}

The incentives for IVR call completion were another dislike, particularly for rural county coordinators:

\begin{quote}
I don’t think that that 25 cents is helpful for motivating people to continue to get the call.

What’s with the incentive? That’s kind of really, make you feel a little worthless.
\end{quote}

Other participants were more open to the idea:

\begin{quote}
Anything that’s an incentive that would give people the extra motivation to want to do it, I think it’s a good idea...it’s not much, but it gives you that sense of, “I made it. I got a quarter, I got 50 cents.” It’s not much, but it’s that knowing that something is in [it at] the end for you.
\end{quote}
Facilitators of and Barriers to Participating in PA or the IVR Intervention

Chronic disease prevention and management was an important motivator for participation in PA or the IVR intervention:

I think depending on where people are in their lives, being physically active might be motivated by so saying, “Hey, this disease process can be kept at bay or managed or maybe even prevented if you exercise.”

Social support was also key to encouraging PA initiation and maintenance, especially once the DIAL intervention ended:

Yes, I do think that they will start or to continue to exercise if they have a friend or a buddy to walk with or whatever. I think that that is important to have someone to exercise with.

I think that the interpersonal aspect of it will be really important. Although it’s not a person, when that connection and accountability with the phone system is removed, I think it will be really important to have that from another source, and hopefully other participants or family members of the participants.

As for barriers to engaging in PA and completing the IVR calls, the participants stressed the lack of time and competing interests:

There are many, many days where I don’t want to do any physical activity. I would say actually most days. It's not because it's tedious. It's because I have so many things to do, and I keep thinking, “Wow. I got to spend that hour doing this.”

Technical Difficulties

The participants described experiencing some initial technical difficulties with the IVR calls, such as receiving calls at incorrect times or with system error messages. The programming decision to skip calls on holidays also seemed to cause some confusion and was changed as a result:

During the 4th of July holidays, I didn’t receive any calls at all that weekend.

Finally, the participants learned to take their time entering the responses during the IVR calls:

If you trying to speed it up and hurry up, you know you going to press two, nuh-uh. it's going to hit you with an error.

Quality of the Calls

The participants generally indicated that the quality of the call is good. In fact, rural county coordinators had previously given the Amazon Polly narration a favorable review at a focus group. During usability testing, several participants distinguished the female voice options as less monotone, robotish than certain boring male voice options. The pace of the calls received mixed reviews; It really was a good pace for some and a bit too fast for others:

There were times where it felt like it was moving a bit too fast, especially when there were multiple options or the question or the prompt was read or said, and then the answers were said immediately after. I don’t know, sometimes it was rushed through, it felt like.

Suggestions

The participants stated that they preferred to have written user-friendly instruction materials that could be used during the calls:

I think you can give them a little prompt card. Like some of the prompts. Because it’s just different every time, but just a small, little introduction of what to expect.

The participants also suggested that having printed materials of the survey readily available to community participants would be beneficial and crucial for capturing accurate data:

I think it would help to have some sort of printout. Just a scale that says, “This is what one means, and this is what 100 means.” Because again, and maybe I just was doing too many things sometimes. For me to remember what that scale was. If I had a call that was coming through, of course I would not click over, but that’s a distraction for me, again. So if I'm in the middle of a question, I'm like, “Oh my gosh. Am I supposed to pick 1 or 100.” I think a printout scale or something in front of me probably would have reminded me, because again, some days I’m putting one, some days I'm putting 99, and that's not what I meant. But there was no way for me to go back and erase my answer, to my knowledge. I don’t know if I missed that in the training, but I just thought, “Oops.” But that was just my short term memory, and knowing I needed to complete the call.

In addition to the advantage of obtaining accurate data, the participants stated that having printed materials would also mitigate noncompliance from frustration:

Anything that we could get to assist would always be helpful. So, if we could come up with something, some type of visual aid for the older generation, then that will be great. I'm sure everybody can work a phone, but you want to make sure that they’re not getting confused. Because once they get confused, confusion causes discouraged sometimes. So, you don't want to get them confused. So, yeah, if we could come up with some type of handout that would be great.

The participants suggested that step goals should be modest and community participants should be reminded to gradually increase their steps during the study:

Take baby steps. I think if you want to see somebody make it to that 10,000 goal marker, try to start small, like see where they're at and see what is an average for the participant and then work your way up from there. And ultimately, it may be just too hard for somebody’s daily schedule to meet that 10,000 goal step without making significant changes to their daily routines. But I think if you take baby steps, then for a good bit of the participants, you may not got to do it for everybody, but at least a certain group of the

https://humanfactors.jmir.org/2022/1/e29494
participants, and you’re able to get some changes, like an increase in daily steps from them, they’re all meeting the goals, then I think that could be a good motivating factor.

It did tell me, try to add a 500 steps for the next time, but I feel like it should be more personalized like, “Okay, you got 2000, tomorrow let’s try to get 2,500.” And then when they call the next day, if they have the 2,500 be like, “Great, you met the goal. Do you think you could add another 500?” I feel like that’s how it should be. I don’t feel like the bar should right off the top be 10,000 because that’s a lot for some people.

The participants also provided solutions for the previously mentioned issues with incentives for call completion (eg, substituting nonmonetary incentives and a point system):

I can tell you that getting 25 cents for each phone call was not motivating at all. Yeah. I don’t mean to be blunt, but it wasn’t. It wasn’t motivating. How much I earned at the end of each phone call, it just didn’t motivate me. Now, if I earned points for each phone call, and I could redeem those points in some sort of physical activity, online store.

I mean, the value may be still $3.75, but with let’s just say 375 points may be like a gift. It may be a pedometer. It may be a little lunch tote, or it may be something else. People can use a tote. They can use a pedometer. They can use even a cup or a mug, or a water bottle if they’re exercising. But what can you do with $3.75?

Finally, the participants discussed the tips provided to increase the number of steps and suggested that they be personalized to the Deep South rural community:

For instance, tell them like get up during commercial break and walk around your coffee table twice. That’s feasible. It’s within reach and it doesn’t take a lot of effort to go out. Because think about these people that don’t have parks nearby. We’re telling them to go to the nearest park. Well, there is no nearest park. Tell them, okay, well walk around your house two times or walk to the mailbox twice or for instance what’s something else. Go three mailboxes down and come back. Something that people can be like, “Oh yeah, I can do that. I never thought to do that.”

Discussion

Summary

Innovative IVR systems hold the potential to overcome barriers to achieving the recommended levels of PA in the rural Black Belt region of Alabama [3]. However, no previous research has examined IVR systems in rural contexts to increase PA levels. We developed an IVR system in an iterative manner based on feedback from earlier pilot studies, focus groups, and the current usability testing with key stakeholders (both community members and local county coordinators with UAB O’Neal Comprehensive Cancer Center Community Outreach and Engagement Office). The resultant system was characterized by high usability and is currently being tested for efficacy in an RCT.

Principal Findings and Resultant IVR System

The IVR system received a numerical usability score of 81—equating to an excellent usability score. The sequential explanatory mixed methods design we adopted helped us identify several opportunities for improvement through the qualitative interviews. After usability testing (qualitative interviews), we implemented several improvements into the system. First, we modified our messaging when the participants failed to reach their goals to sound less negative or stern. We implemented graceful handling of wrong key presses by participants; instead of informing them that they had pressed a wrong key, we reworded to say that the system could not understand. We implemented a detailed orientation session procedure in which the IVR system was oriented and printed materials were made available. Instead of directly pushing the participants toward 10,000 steps, following participant feedback, we implemented incremental goals of 250 steps per week. We reworded our reward system to use the word points instead of cents to emphasize the gamification of IVR adherence versus financial transactions. Finally, we also added several more PA tips as suggested by our interview participants.

As a means to further test the system before the commencement of the RCT, the core group of researchers working on this study met to discuss whether further formal usability testing was required. As most reported issues pertained to wording or content, it was decided that no further usability testing was needed. However, the core group of researchers were listed as pilot users of the IVR system. These pilot users were scheduled approximately 2-3 weeks ahead of the actual RCT participants. These researchers actively tested the system daily and reported to the development team on any issue found. This enabled the development team to aggressively address the issues before any RCT study participants encountered them. Some example issues identified and fixed using this approach included problems when the participants moved from one phase (daily calls) to another (biweekly calls) and nonavailability of new PA strategies to suggest to participants. Our 2 weeks ahead approach enabled us to resolve these issues before any real participants encountered them while avoiding lapses in time that would delay the project.

Our final product is a comprehensive IVR system with cutting-edge capabilities such as streamlined calls, smart dropped calls handling, and assignable voice gender. Future research should examine the added value of such features and their impact on this promising technology.

Final IVR System Design

The feedback during, before, and after the usability testing was used to iteratively refine the IVR system. The resultant system, which is now being used in the RCT, is described as follows:

1. The system is designed to handle complex call schedules involving different types of calls during different phases of the intervention and the randomization group.
2. The system can receive incoming calls and smartly place outgoing calls only if the participant has a pending incomplete call.

3. To protect the privacy of the participants, they are identified using their phone number and a personal identification number (PIN). When the participants use their registered phone, only a PIN is required. When the participants use a phone other than their registered phone, both the registered phone number and PIN are required. This achieves a balance between user experience and security.

4. New participants are registered on a web portal by a study manager who retrieves the unique PIN for the user. The study manager is then able to print the PIN and other instructional materials in an educational binder for the participants.

5. A comprehensive missed call policy has been implemented, with the system retrying the call after 30 minutes. Again, if there is no response, the call is marked as incomplete and can be completed the next day.

6. A smart dropped-call policy has also been implemented, wherein if a participant drops midway through a call and the user connects again within a preset time limit, the participant is able to continue from the last question they answered.

7. One of the most significant aspects affecting the usability of IVR systems is the information navigation time [7]. We have essentially eliminated the navigation time by streaming the content for calls in multiple ways. First, calls are not placed unless there is a pending survey to be completed. Second, when users have multiple surveys due, the system combines all the surveys and offers them in a sequence. Finally, if the participant has any pending surveys as a result of missed calls in the previous 2 days, the system offers the missed surveys in sequence.

8. Many IVR systems require a significant amount of time because of the confirmation messages, such as “You pressed 6, press 1 if this is right or press 2 to change.” These confirmation messages are necessary as it is easy for a participant to accidentally mistype a number; however, these confirmation messages almost double the call time. To overcome this, during the orientation session, we educate the users on pressing * anytime during the call to edit the last response.

9. The participants can call the IVR system anytime and change their preferred call receiving time.

10. To maximize information relevance [7], the system is programmed to be able to look up the participant’s previous step goals, PA self-efficacy, enjoyment, social support, and outcome expectancies and use those values as a part of the conversation—thereby leading to high relevance.

11. To maximize information capacity [7], a bank of PA-increasing strategies has been created, with new strategies being revealed on a weekly basis. Similarly, a bank of greeting messages has also been made available. Through these mechanisms, despite the daily calls during the first 3 months, the users would find a variety of content being delivered.

12. At the end of each call, the system announces the reward points earned by the participant, which can be redeemed for actual monetary incentives. Before the usability testing, we directly referred to points as cents. However, we learned from usability testing that the participants felt that 25 cents per call made them feel that their time was worthless. Thus, we reworded our call content to award points rather than cents.

13. The gender of the voice narration in the call can be set to match the gender of the participant, to the opposite gender, or to be random.

Strengths and Limitations
This study had a few limitations. First, usability testing was conducted during the COVID-19 pandemic; thus, for participant safety, all surveys and interviews were conducted remotely, and participation was limited to community health advisors and staff. Although these community health advisors live and work in the same rural counties of the Black Belt region of Alabama as the future participants, it is possible that they do not accurately represent the demographics of the participants (eg, education levels) who would participate in the RCT study. In addition, our demographics includes predominantly female and non-Hispanic or Latino populations.

However, this opportunity allowed rural county coordinators to gain familiarity and comfort with the inner workings of the IVR system before spearheading its dissemination among their own communities. Thus, they will be more prepared to orient participants to the IVR system and field their questions. Moreover, playing such a key role in the development and refinement of this technology likely enhanced the sense of buy-in and ownership among these key stakeholders and gatekeepers to the community and substantially improved the final product.

Conclusions
This study demonstrated that the developed IVR system is usable and has the potential to increase the levels of PA. Study findings provided insight into the participants’ preferred language, narration tones, rewards, and variety of messaging. These insights can be valuable for future studies that seek to develop IVR-based interventions.

Acknowledgments
This study was funded in part by the National Cancer Institute (R01CA233550) and the National Heart, Lung, and Blood Institute training program (T32HL105349).
Authors' Contributions

MT led the development of the interactive voice response system and was thus not involved in the usability testing of the system. DP and SN, who were not part of the interactive voice response system development team, conducted the usability tests and interviews.

Conflicts of Interest

None declared.

References


Abbreviations

- **DIAL**: Deep South IVR-Supported Active Lifestyle
- **IVR**: interactive voice response
- **PA**: physical activity
- **PIN**: personal identification number
- **RCT**: randomized controlled trial
- **SCT**: social cognitive theory
- **UAB**: University of Alabama at Birmingham

©Mohanraj Thirumalai, Nashira Brown, Soumya Niranjan, Sh’Nese Townsend, Mary Anne Powell, Whitney Neal, Erica Schleicher, Venkatadri Raparla, Robert Oster, Wendy Demark-Wahnefried, Dori Pekmezi. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 04.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.