Journal of Participatory Medicine

Advancing the understanding and practice of participatory medicine among health care professionals and patients
Volume 15 (2023)  ISSN: 2152-7202  Editors: Susan Woods

Contents

Original Papers

Co-design of the Transgender Health Information Resource: Web-Based Participatory Design (e38078)
Brad Morse, Andrey Soares, Kate Yell, Kristen DeSanto, Marvyn Allen, Brooke Holliman, Rita Lee, Bethany Kwan, Lisa Schilling......................... 2

Enhancing Exposure Treatment for Youths With Chronic Pain: Co-design and Qualitative Approach (e41292)
Lea Schemer, Courtney Hess, Amanda Van Orden, Kathryn Birnie, Lauren Harrison, Julia Glombiewski, Laura Simonsa............................. 17

Acceptability of Automated Robotic Clinical Breast Examination: Survey Study (e42704)
George Jenkinson, Natasha Houghton, Nejra van Zalk, Jo Waller, Fernando Bello, Antonia Tzemanaki.................................................. 35

Development of a Family-Centered Communication Tool for Kidney Health in Premature Infants: Qualitative Focus Group Study Using Human-Centered Design Methodology (e45316)
Michelle Starr, Samantha Wallace, Courtney Moore, Brandon Cockrum, Bridget Hawryluk, Aaron Carroll, William Bennett Jr...................... 51

Review

Examining Patient Engagement in Chatbot Development Approaches for Healthy Lifestyle and Mental Wellness Interventions: Scoping Review (e45772)
Chikku Sadasivan, Christofer Cruz, Naomi Dolgoy, Ashley Hyde, Sandra Campbell, Margaret McNeely, Eleni Strouila, Puneeta Tandon.................. 63
Co-design of the Transgender Health Information Resource: Web-Based Participatory Design

Brad Morse, MA, PhD; Andrey Saurés, PhD; Kate Ytell, MPH; Kristen DeSanto, MSLS; Marvyn Allen, MPA; Brooke Dorsey Holliman, PhD; Rita S Lee, MD; Bethany M Kwan, PhD; Lisa M Schilling, MSPH, MD

1Division of General Internal Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
2Data Science to Patient Value Initiative, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
3Elevance Health, Denver, CO, United States
4Strauss Health Sciences Library, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
5One Colorado, Denver, CO, United States
6Department of Family Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
7Adult and Child Center for Outcomes Research and Delivery Science, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
8Department of Emergency Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Corresponding Author:
Brad Morse, MA, PhD
Division of General Internal Medicine
University of Colorado Anschutz Medical Campus
1890 North Revere Court | Mailstop F443
Aurora, CO, 80045
United States
Phone: 1 3037245138
Email: brad.morse@cuanschutz.edu

Abstract

Background: There is an urgent and unmet need for accessible and credible health information within the transgender and gender-diverse (TGD) community. Currently, TGD individuals often seek and must find relevant resources by vetting social media posts. A resource that provides accessible and credible health-related resources and content via a mobile phone app may have a positive impact on and support the TGD population.

Objective: COVID-19 stay-at-home orders forced a shift in the methods used in participatory design. In this paper, we aimed to describe the web-based participatory methods used to develop the Transgender Health Information Resource. We also described and characterized the web-based engagement that occurred during a single session of the overall design process.

Methods: We planned and conducted web-based design sessions to replace the proposed in-person sessions. We used web-based collaborative tools, including Zoom (Zoom Video Communications), Mural (Mural), REDCap (Research Electronic Data Capture; Vanderbilt University), and Justinmind (Justinmind), to engage the participants in the design process. Zoom was used as an integrated platform for design activities. Mural was used to perform exercises, such as free listing, brainstorming, and grouping. REDCap allowed us to collect survey responses. Justinmind was used to create prototypes that were shared and discussed via Zoom. Recruitment was led by one of our community partners, One Colorado, who used private Facebook groups in which web-based flyers were dispersed. The design process took place in several workshops over a period of 10 months. We described and characterized engagement during a single design session by tracking the number of influential interactions among participants. We defined an influential interaction as communication, either verbal or web-based content manipulation, that advanced the design process.

Results: We presented data from a single design session that lasted 1 hour and 48 minutes and included 4 participants. During the session, there were 301 influential interactions, consisting of 79 verbal comments and 222 web-based content manipulations.

Conclusions: Web-based participatory design can elicit input and decisions from participants to develop a health information resource, such as a mobile app user interface. Overall, participants were highly engaged. This approach maintained the benefits and fidelity of traditional in-person design sessions, mitigated deficits, and exploited the previously unconsidered benefits of web-based methods, such as enhancing the ability to participate for those who live far from academic institutions. The web-based approach to participatory design was an efficient and feasible methodological design approach.
Introduction

Transgender and Gender-Diverse Health and Medical Information Needs

Transgender and gender-diverse (TGD) individuals (defined as people whose current gender is different from that assigned at birth, including, but not limited to, nonbinary, queer, and gender nonconforming people, hereafter shortened to TGD) face health disparities including high degrees of stigma and discrimination from providers and health care systems [1-4]. Three-fourths of TGD individuals report negative experiences with the health care system [5]. TGD individuals report difficulties finding and accessing TGD-competent health care professionals [6-10], securing insurance coverage for their health care needs [11-13], and finding health care professionals who are sensitive to the needs of the TGD population [14]. Transgender individuals experience stigma and discrimination across the social determinants of health, including bullying in schools, lack of stable income, and quality housing [15-17]. Moreover, TGD individuals often have to manage chronic stress owing to traumatic experiences over their life course [18,19].

Owing to stigma and discrimination when seeking health care, TGD community members often turn to health and medical information on the web [20,21]. Documented examples of insensitive health care include gender insensitivity in which individuals were misgendered (using “he” when a “they” pronoun was requested) or forced care where some patients felt they were forced to do unnecessary examinations or dismissed as “psych cases” [22]. A growing body of literature focuses on the TGD community and their health information–seeking behavior on the web [20,23,24]. A study found that gender transition mental health message boards are popular, especially on the Tumblr platform [25]. In another study, younger TGD individuals used various web-based platforms to explore transgender, nonbinary, and gender-diverse identities and to find support networks [26]. There is evidence that social media is a key resource relied upon by the TGD community to obtain health and medical information [27]. The affordances of social media provide a network for peer-to-peer, emotional, appraisal, and informational support [27].

Considering the known difficulties in seeking and determining the credibility of web-based transgender health information and the current lack of transgender-specific materials on the web [20,24], we aimed to create a health information resource to support the TGD population. For this project, credible information was defined as information created or disseminated by clinicians and organizations, such as the Trevor Project [28], with expert knowledge about care specific to the needs of transgender individuals, and reputable sources of health information such as MEDLINEPlus [29]. Published literature highlights the lack of credible web-based information resources dedicated to the needs of TGD individuals [20,30,31]. Digital tools might be important for TGD health self-management, but they are currently underutilized [32].

To address this need, we developed the Transgender Health Information Resource (TGHIR). We chose a mobile app (Android and iOS mobile operating system) user interface as the optimal method of delivery because of mobile phone ownership—85% of Americans own a smartphone [33]—and mobile phones support anywhere and anytime access to information.

Participatory Design and Web-Based Participatory Design

Participatory design has proven to be successful in designing mobile health resources. The benefits of this approach include engaging co-designers (henceforth, participants) selected to be representative of the community of intended end users to thoroughly explore and prioritize target audience needs [34-36]. Researchers and participants can collaborate and design [37] interfaces that make information accessible. Representation in the design process is helpful because researchers often do not understand how others are affected by technology performance [38].

Leveraging the tacit knowledge and lived experiences [39,40] of individuals from a community helps to understand how everyday tasks are conceptualized, approached, and completed. This process is typically made easier through in-person design sessions. The benefits of in-person collaboration include rapport building [41], a sense of ownership from participants who co-design the system [42], shared values on which design facilitators can build energy within the participatory design methodology [39], as well as perceiving subtle cues of interpersonal communications such as facial expressions and body language [40]. During in-person design sessions, brainstorming and drawing activities are conducted using tangible tools, including butcher paper (a type of heavy paper hung on walls to collect and record ideas) and sticky notes for rearranging the linkages, groupings, and prioritization of specific ideas, allowing for rapid iteration. These activities, and their necessary physical tools, support collaboration and inspire meaningful dialogue between participants and researchers [43].

Approaches to evaluating participatory design have focused on the processes deployed, effects on designers, and outcomes such as satisfaction and empowerment [44,45]. Evaluation of the design process and decision-making have included the collection of qualitative data from end users and system developers to determine the effectiveness of the decision-making [46]. When the focus was on the effect of the design process on participants, evaluators concentrated on participant experiences through interviews [44]. Outcomes, such as participant gains, can also be evaluated, by assessing participant experiences through interviews or surveys to measure how the design product
benefited them and if that benefit lasted [47]. The methodological discipline of participatory design is based on meaningful communication between participants and researchers and allows decision-making by participants, such as how health information is accessed and displayed.

The TGHIR design process coincided with the early months of the COVID-19 pandemic, which forced a transition from in-person to web-based design sessions. Although remote collaboration for design-based user experience interviews was becoming more common [48,49], the use of a suite of web-based tools to support web-based design collaboration has been less documented, especially in the TGD community [50]. Early participatory design studies [50-52] focused on how communication technologies such as email and websites could organize the web-based design process. The effective use of these communication tools, in combination with a shared web-based space hosted on the internet where design could happen synchronously or asynchronously, led to the advancement of how participatory design could be implemented. The addition of shared web-based creative spaces supported participation among remote team members while requiring fewer resources.

Web-based participatory design to support community-driven development of products on an asynchronous discussion forum has proven to be successful. Research conducted by Hess and Pipek [53] indicated that engaging web-based communities to support community-driven development of consumer software is possible, especially if the work is intrinsically fun. The authors found that participants on the web could contribute to the design process of a software system. However, the project began to feel unpaid by some members. A power balance between participants on the web and professional designers was observed and influenced the decisions made for the system. The findings of this study suggested that the responsibilities of participants on the web should be limited to distinct use cases so the development process is not dominated by the most engaged volunteers who might have affordances, such as levels of experience or more time to participate, which may allow greater influence on design decisions. Our proposed web-based participatory design addressed some of these challenges.

Participatory design has been used by other research teams to support the design of health information resources [54-59]. This project was potentially the first to engage TGD participants exclusively in a web-based participatory design process. The web-based method of engagement might safeguard privacy and safety and allow greater involvement of TGD individuals in research. The overall objective of this paper was to describe our web-based participatory methodology and engagement evaluation for design session 1.

Methods

Overview

The setting was the University of Colorado Anschutz Medical Campus, an urban academic medical center, in collaboration with the University of Colorado Integrated Transgender Clinic, and One Colorado, the state’s leading advocacy group for LGBTQ (lesbian, gay, bisexual, transgender, and queer) persons. The design and development of the TGHIR were guided by a participatory process to ensure that the final TGHIR design would serve the health needs and goals of the TGD community. The process involved deploying a series of iterative methods to explore the use context and needs of end users. We described and characterized web-based methods and engagement, including qualitative insights from participants on the design of the TGHIR and the number of influential verbal and web-based participant interactions.

Recruitment

There were 3 groups involved in the design of the TGHIR: researchers, advisers, and participants. Researchers were responsible for facilitating the design sessions and implementing the decisions made in partnership with the advisers and participants as well as the agile [60] development of the resource. Advisers were partners [61,62] in the design process and provided feedback on participant engagement strategies, insights from design sessions, and the development process. Participants were responsible for generating ideas and making decisions on how the TGHIR should be designed, features to include, and wireframing.

Participants had to be aged 18 years and meet one of the following two inclusion criteria: (1) self-identified as TGD at any point in their journey or (2) be parents or guardians of a TGD youth.

Recruitment took place during the COVID-19 stay-at-home orders from the State of Colorado. Strategies for recruitment included posting web-based flyers in TGD Facebook groups and delivery through listservs. The outreach was led by a local TGD community partner, One Colorado [63]. We conducted a thorough eligibility call over the Zoom videoconferencing platform (Zoom Video Communications) with each participant to assess their interest in transgender health and how they would like to contribute, which allowed the researcher to confirm the intent of their interest and that the participants could connect to Zoom [64]. Community Advisory Board members, known as advisers, were nominated, after funding for the project was obtained, by partners and allies in the transgender community, One Colorado, and the Integrated Clinic at the CU Anschutz Medical Campus. The process of adviser selection included a nomination phase and an interview phase to assess whether the nominee’s goals and motivation fit with the objectives of the project. All nominated individuals were onboarded. Community Advisory Board membership included 8 TGD individuals, 3 parents of TGD adolescents, 1 advocate from One Colorado, 4 health care providers who served the TGD population, 1 library scientist, and 4 research staff, totaling 18 members (3 advisers were in 2 of the categories described above). Advisers attended a 4-hour in-person kickoff meeting and eight 1-hour web-based meetings. The topics of the 8 web-based meetings included purpose-to-practice exercise, identifying credible health information, asynchronous work group planning, a discussion on the Black Lives Matter movement in the summer of 2020, a review of health care provider resources, focus group and design session data discussion, TGHIR app demo feedback,
usability testing data discussion, and a celebration to acknowledge what we achieved together. Each adviser was compensated US $50 per hour for attending the meetings.

**Participatory Design Sessions**

A web-based design approach required adapting in-person participatory design approaches to understand the targeted end users, the tasks end users were attempting to complete, and the environment in which participants completed the tasks [59]. The three stages of the TGHIR participatory design approach were adapted from Spinuzzi [65] and included (1) initial exploration of end-user needs, (2) discovery processes of prioritization and ideation on potential outcomes, and (3) prototyping. As a research team, we integrated these stages into a larger design process to access the opinions and experiences of participants on behalf of the targeted end users. A total of 4 distinct design sessions were conducted as part of the TGHIR design process (Table 1). We described the methodology implemented in these sessions, including how the methods were adapted to web-based interactions and the degree of participation in the “Ladder of Participation” column [61,62], a theoretical construct that described the range of co-designer participation from low to high.

Several web-based tools were used for web-based participatory design session implementation, including Zoom [64], Mural (Mural) [66], REDCap (Research Electronic Data Capture; Vanderbilt University) [67], and Justinmind (Justinmind) [68]. Zoom was used primarily to allow all participants to meet, with audio and video capabilities using a computer or tablet and computing functionality in a common web-based space. Mural was used as the collaborative workspace for the design sessions; participants were asked to create free accounts and were then linked to collaborate in a Mural workspace. REDCap was used for web-based surveys and Justinmind software was used for wireframing and prototyping.

### Table 1. Four design sessions.

<table>
<thead>
<tr>
<th>Design session and goals (research participants)</th>
<th>Participatory design stage adapted from Spinuzzi [65]</th>
<th>Ladder of participation</th>
<th>Web-based tools</th>
<th>Planned in-person tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 1: Exploration of potential features (n=4)</td>
<td>Stage 1: initial exploration of end-user needs</td>
<td>Consultants: participant feedback on features</td>
<td>Mural and Zoom</td>
<td>Butcher paper, sticky notes, markers, and bullseye visualization</td>
</tr>
<tr>
<td>Session 2: Feature prioritization (n=22)</td>
<td>Stage 2: prioritization (Kano) and ideation on potential outcomes</td>
<td>Consultants or partnership: participant feedback via Kano and decisions on prioritization</td>
<td>Mural, Zoom, and REDCap¹</td>
<td>Paper survey, butcher paper, sticky notes, markers, and bullseye visualization</td>
</tr>
<tr>
<td>Session 3: Iterative prototyping</td>
<td>Stage 3: prototyping</td>
<td>Partnership: participants decide how the resource displays each feature</td>
<td>Mural, Zoom, and Justinmind</td>
<td>Butter paper, sticky notes, markers, PowerPoint to create first draft screen visuals, prototyping software</td>
</tr>
<tr>
<td>Session 4: Usability testing (n=2)</td>
<td>Heuristic evaluation: cognitive walkthrough</td>
<td>N/A²</td>
<td>Zoom, REDCap, and the TGHIR²</td>
<td>The TGHIR, paper, and pencil</td>
</tr>
</tbody>
</table>

¹REDCap: Research Electronic Data Capture.
²N/A: not applicable.
³TGHIR: Transgender Health Information Resource.

**Design Session 1: Exploration of Potential Features**

We described and characterized engagement in this session and provided findings in the results. Overall, design session 1 included 4 participants who identified as transgender (2/4, 50%) and nonbinary (2/4, 50%). The races reported by participants included American Indian (1/4, 25%), multiple races (1/4, 25%), and White (2/4, 50%). Age ranges were as follows: 20 to 29 years (1/4, 25%) and 30 to 39 years (3/4, 75%). Participants lived in rural areas (1/4, 25%) and urban locations (3/4, 75%).

Design session 1 was exploratory, informed by stage 1 of participatory design approach by Spinuzzi [65] and facilitated by the first author, a member of the research team. The focus was on gathering information from participants about the mobile resources end users in our target audience liked using and what features made their experience with the resource enjoyable. Participants developed personas through an exercise in which they ascribed feelings, values, and behaviors, resulting in potential use cases. An example of a persona statement is provided in Figure 1.

Mural (Figure 2), a web-based collaborative platform, allowed us to use the electronic counterparts of butcher paper, sticky notes, electronic pens with assorted color ink, and distinct colors and sizes of fonts.

We held a 10-minute Mural training session to optimize the time we had with the participants. Owing to the exploratory nature of design session 1, multiple brainstorming activities were conducted. Typically, participants wrote on sticky notes and organized their ideas on a whiteboard. Mural allowed for this functionality. In this session, we asked participants a series of questions about their mobile resource preferences, including, (1) What mobile apps do you like and why, (2) What mobile apps do you dislike and why, (3) What mobile apps do you use often and what makes them reusable, (4) What mobile apps have you stopped using and why, (5) What makes information on a mobile app credible, and (6) What websites do you use to
Design Session 2: Feature Prioritization

Design session 2 focused on discovering and prioritizing resource features identified in the first design session using stage 2, the discovery processes, of the participatory design approach [65]. The prioritization of features was organized using a Kano model of customer satisfaction [69-72]. This method assessed participants’ opinions regarding a feature being implemented [69]. The transition to web-based implementation required our design team to collect the Kano survey of customer satisfaction through REDCap instead of on paper and in-person. In this exercise, we used Mural to display the features generated by participants in design session 1 so the feature could be prioritized. The participants went through the survey as a group but responded individually.

After the Kano survey was completed, a second round of placing features and categories on a priority bullseye visualization (Figure 3) was conducted as a design exercise, furthering the discussion on what features were the most desirable to end users and should therefore be developed first. The design session 2 bullseye exercise was compared with how features were prioritized using the Kano survey. In addition, this exercise allowed participants to group the features into categories, such as in-person card sort, by grouping different sticky notes that referred to similar or the same type of features [73,74]. The last activity in design session 2 was the initial prototyping of the home menu and health resource preview, which was performed in Mural.
Design Session 3: Iterative Prototyping

Design session 3.1 and design session 3.2 focused on wireframing, a process in which a sketch is made of what a product, in this case, a user interface, may look like. The wireframe was used as a starting point for the design work [75] and prototyping of the health information resource as an intentional and planned health information-seeking experience in stage 3 of the participatory design approach [65]. This work was performed using the Justinmind tool, a program that allows designers to rapidly create interfaces and modify them in real time. Using Justinmind, we created wireframes and asked participants for input on the design and implied functionality using Zoom. Participants first responded to the design elements of the baseline design created by the researchers. In traditional in-person sessions, the pencil and paper design methodology allowed for a quick iteration of the initial design during an in-person session. In this web-based design session, the use of a baseline design allowed the researcher to engage more quickly and in greater detail about the interface because time was not used to set up the initial design. Design session 3.3 was held with one participant to collect insights into aesthetics, or look and feel, of the TGHIR.

Design Session 4: Usability Goal-Oriented

The fourth and final design session occurred after the TGHIR was built and consisted of a cognitive walkthrough heuristic evaluation, in which an experienced mobile health researcher and a behavioral scientist were asked to perform navigationally based tasks, assuming that most tasks embedded in the TGHIR would be goal-orientated [76], for example, find and like a resource. Participants installed the resource on their phones for an authentic experience.

In the usability testing sessions, Zoom, REDCap, and TGHIR were used. In our initial plans, this interaction would have been in-person and the participant would have had the TGHIR in
their hands, as we observed their use of the resource and recorded the necessary usability data. Instead, design team members observed the participants through a Zoom connection. Usability evaluation tasks (Table 2) were described and completed by the participants [77,78]. These tasks were prioritized because they were associated with the features participants identified as important for the TGHIR to access credible health information.

Table 2. Evaluation tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Action to be completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Create account and view consent and privacy language</td>
</tr>
<tr>
<td>2</td>
<td>Select preferred health information categories</td>
</tr>
<tr>
<td>3</td>
<td>Find a specific health information item using the category cards</td>
</tr>
<tr>
<td>4</td>
<td>Find and use the filter to narrow the resources to a relevant informational item</td>
</tr>
<tr>
<td>5</td>
<td>Like a resource</td>
</tr>
<tr>
<td>6</td>
<td>Bookmark a resource</td>
</tr>
<tr>
<td>7</td>
<td>Locate and use the search function to find a specific health information resource</td>
</tr>
<tr>
<td>8</td>
<td>Send a message to the developers</td>
</tr>
<tr>
<td>9</td>
<td>Share a new resource for the community with the developers</td>
</tr>
<tr>
<td>10</td>
<td>Find the most liked health information resource</td>
</tr>
</tbody>
</table>

Data Analysis
We analyzed the video recording from design session 1 and the artifact created in the Mural collaborative space to describe and characterize web-based engagement. Design session 1 was recorded using the Zoom tool and transcribed by a professional transcriptionist. A research team member performed rapid analysis [79] of the transcriptions to quickly identify key points and comments that reflected participant engagement and major design decisions. Exemplar quotes were provided to highlight the impact of comments on the final design and the resulting features developed in the resource. Quantitative data included length of each design session exercise, number of verbal comments by all participants, total verbal comments, number of times the Zoom camera was turned off by participants, number of Mural interactions (creation or manipulation of digital Mural content) by participants, and total Mural interactions by design exercise. For a verbal comment or Mural interaction to be counted, the comment or interaction had to be considered influential in advancing the design process. Comments or interactions that expressed confirmation or agreement were not considered influential.

Ethics Approval
Informed verbal consent was obtained from all participants. Each participant was paid US $50 for each of the 1.5-hour design sessions they attended. Project approval was obtained from the University of Colorado’s Human Research Ethics Committee, Colorado Multiple Institutional Board (COMIRB# 19-1562).

Results
Recruitment
Eligibility screening was performed using Zoom. A total of 41 individuals were screened for eligibility; of these, 27 (66%) were eligible and available to participate in at least one of the design sessions. Mural relied on well-established computing conventions to add, edit, and delete content and was, therefore, familiar to the participants who quickly mastered the skills to work alongside the research staff in the collaborative space.

Design Session 1
In design session 1, participants identified the following specific health information for inclusion in the resource: affirming care (supportive care for the TGD community [7]), affordable medical options, information on transition, information on successful transitions (transitions were associated with the period during which a person began to live according to their gender identity rather than the gender they were thought to be at birth [80]), family resources, community support, and medically verified information.

Responses to the questions addressing participants’ mobile app preferences are presented in Table 3.

We provided the number of verbal comments and web-based interactions in design session 1 to show the volume of participant engagement in Table 4. Design session 1 was 119-minutes long; there was an average of 1.87 web-based points of engagement every minute. The researchers turned off their cameras to focus on the participants. Two participants experienced bandwidth trouble during the first exercise and chose to turn off their cameras to contribute to the web-based participatory design process. These 2 participants did not provide any other reason for their cameras being turned off and both continued to contribute despite the bandwidth problems they experienced.

Additional qualitative evidence for the effectiveness of the web-based participatory design process is presented in Table 5. Participant insights from the first design session remained prominent throughout the design process and directly affected the course of the design work. Participant’s verbal interactions were evident in the final TGHIR features.
Table 3. Questions addressing participant’s mobile resource feature preferences.

<table>
<thead>
<tr>
<th>Question</th>
<th>Responses</th>
</tr>
</thead>
</table>
| What mobile apps do you like and why? | • Gmail: simple design, easy to access, can limit notifications  
• Facebook: easy to refresh  
• Uber  
• Reddit: the ability to search by subcategories  
• Spotify  
• Evernote: does a fair job of talk to text  
• Merlin Bird ID: simple, easy to use, and regularly updated  
• Mint: great UX\(^a\) and UI\(^b\) and never buggy  
• Native Land: great multisource database and simple UI and UX  
• AllTrails: surprisingly prefer using mobile over web app because of UI and UX |
| What mobile apps do you dislike and why? | • Whole Foods: not intuitive  
• Snapchat: too many notifications  
• Your Turn: advertisements interrupting experience  
• Tabletopia: optimized for PC and Tablet not mobile  
• Spectrum Mobile: lack of control with certain items  
• C25K: too many options  
• Stitcher: too many options  
• Apple Notes app: clunky and difficult to change formats  
• Google sheets: difficult to use  
• Apple Maps: location and information often totally inaccurate |
| What mobile apps do you use often and what makes them reusable? | • Hearthstone: rewards frequent use  
• Facebook: allows me to keep connected  
• Reddit: plenty of media available  
• Discord: able to access through multiple mediums  
• Smarthub: saves relevant data for later  
• Keep Notes: does exactly what it needs  
• Instagram: addictive, friends, distraction  
• Stitcher: do not like the interface, I just like the content, streaming, and downloading  
• NYTimes: good usability  
• Gmail: necessity and like UX better than other mail apps  
• Spotify: seamless desktop to mobile transition |
| Reasons for not using an app | • Excessive or random crashes  
• Too many ads  
• Excessive notifications  
• Difficult to navigate; no search function  
• Too many options  
• Need to create an account when not necessary  
• Ugly or outdated  
• Uses too much power or memory on phone  
• Busy user interface |
| What makes information on a mobile app credible? | • Citing sources  
• Current information  
• Customer service  
• Association with credible groups  
• Not asking to rate the app  
• Easy to navigate  
• Inclusive language  
• Transparency on who developed app |
| What websites do you use to find health information? | • Private Facebook groups  
• Forums like Reddit  
• WebMD  
• Denver Health  
• PFLAG  
• One Colorado  
• Queer Asterisk |

\(^a\)UI: user interface.  
\(^b\)UX: user experience.
Table 4. Design session 1 engagement evaluation (total interactions, N=301).

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Exercise time (length of discussion)</th>
<th>Verbal comments by participants, n</th>
<th>Verbal comments per exercise, n</th>
<th>Web-based interactions by participants, n</th>
<th>Web-based interactions per exercise, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mural exercise</td>
<td></td>
<td>A² B³ C³ D³</td>
<td>A³ B³ C³ D³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What apps do you like</td>
<td>6:10-14:50</td>
<td>2 2 0 1 5</td>
<td>8 10 3 7 28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What apps do you use</td>
<td>14:51-22:29</td>
<td>1 1 3 3 8</td>
<td>11 10 7 10 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credible information</td>
<td>22:30-30:30</td>
<td>5 3 0 1 9</td>
<td>5 5 4 4 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGD² information</td>
<td>30:31-39:05</td>
<td>1 2 0 0 3</td>
<td>5 4 4 2 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brainstorm—free listing</td>
<td>39:06-1:00:38</td>
<td>8 7 1 3 19</td>
<td>10 8 9 8 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Card sort</td>
<td>1:00:39-1:13:24</td>
<td>3 5 2 2 12</td>
<td>12 14 10 10 46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bullseye prioritization</td>
<td>1:13:25-1:28:00</td>
<td>7 7 3 3 20</td>
<td>8 12 8 0 28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value proposition generation</td>
<td>1:34:05-1:55:10</td>
<td>2 1 0 0 3</td>
<td>4 4 2 4 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (including breaks)</td>
<td>1:58:55 (including breaks)</td>
<td>29 28 9 13 79</td>
<td>63 67 47 45 222</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

²Participants.
²TGD: transgender and gender-diverse.

Table 5. Exemplar quotes, impact, and design influence.

<table>
<thead>
<tr>
<th>Session Exercise</th>
<th>Result</th>
<th>Impact on design</th>
<th>Influence on design</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3</td>
<td>“I think having a channel in which folks can reach out if there’s an issue with information on the app…message customer support, call customer support or some kind of way that it’s not just, you know, you’re not searching really, you know, a long way through the app or trying to find a human to connect with within even an issue or a question”</td>
<td>Customer support as a priority within the app—visible, easily accessible. Important to make users feel valued and heard.</td>
<td>Contact Us page with the option of sending different types of messages to the TGHIR² development team.</td>
</tr>
<tr>
<td>1.5</td>
<td>“I think it would be good if there is a place for it [the resource] to remember your name but also not to make it too cumbersome should you decide to change it. If it remembers your names also having an option for pronouns, even if it doesn’t pop up anywhere, it’s nice to have that affirmation.”</td>
<td>Function to remember names and customize pronouns within the resource. Demonstrates understanding of target audience.</td>
<td>Providing pronouns was eventually scrapped to avoid collecting data the resource would not use. This is the settings menu based on the comment above. The design attempts to make changing settings easy.</td>
</tr>
<tr>
<td>1.5</td>
<td>“My Verizon [app] can update some things on your account, but otherwise it’s just a giant ad for you to upgrade your system. I think interactivity, there’s a reason why you’ve downloaded an app instead of going to the website.”</td>
<td>Information should be presented in a mobile app that fosters easy access with an intuitive interactive navigation so the app could be used at any time and in any place.</td>
<td>The TGHIR was design and developed for iOS and Android, the 2 most common mobile app operating systems.</td>
</tr>
<tr>
<td>1.6</td>
<td>“I have the [Name], My Chart open. And one of the options is you can personalize it. The personalization is just changing the color for the frame of the app. That seems pointless to me…yeah, it just seems like such a superficial option. I would rather move the icons around on the homepage…there might be something that I like to check more than other people. Moving the menu around would be a nice personalization. But changing the color is not worth it.”</td>
<td>If the resource is going to include customization options, it should go beyond the ability to simply change the colors on the interface, eg, the option to move icons around based on relevancy to the individual user. The orange cards are categories of interest selected by the user.</td>
<td>The Health Resource Cards could be pinned to the top of the page making them easier to return to during repeated use. Once the resource was pinned, the TGHIR app changes the card from blue to orange.</td>
</tr>
</tbody>
</table>

²TGHIR: Transgender Health Information Resource.
Discussion

Principal Findings

We found that using a web-based suite of collaborative tools with participants, we were able to effectively engage in productive discussions and make design decisions for the development of TGHIR. This approach to design work, in our case, was seamless and did not limit participants from engaging and providing meaningful influence on the resulting health information resources. The value of web-based recruitment and design sessions should be underscored because of the relative social safety of TGD participants. This recruitment method may be a strength for engagement and have a positive impact on TGD involvement in research by creating safe web-based spaces for TGD involvement.

We observed consistent engagement throughout design session 1. Upon reflection and analysis of the data, we found that the immediacy of engagement was impressive. It did not take participants much time to get involved and discuss design issues with their fellow participants. The dual interaction of manipulating web-based content in the Mural document and discussing the topic at hand may have led to a more immediate collaboration.

This study shed light on a web-based methodological approach to co-design health information resources within TGD communities. Through their involvement and enthusiasm for the work, participants indicated that a web-based approach to design was appropriate and can be used instead of resource-intensive in-person gatherings. Queer and trans communities have embraced digital technologies in radically affirming ways [81,82] to move beyond the acceptance of unsatisfactory options. The uptake of web-based design, and the necessary digital technologies, in TGD communities is feasible.

Mural interactions indicated that participants were consistently present throughout the design process. With 222 Mural interactions across 4 participants, it was evident that these individuals were highly engaged on the process. Whether this level of engagement can be replicated or generalized to other communities and topics requires further research. The verbal comments provided evidence that the interactions were process-oriented and moved the discussion to inform the development of TGHIR. Furthermore, specific design decisions were made owing to the input of the participants, providing evidence of their importance in the web-based design process and speaking to the effectiveness of web-based participatory design.

Strengths, Limitations, and Future Directions

The strength of this project was the introduction of a method to describe and characterize engagement and interactions during participatory design facilitated by web-based means. The first limitation was that this study evaluated the engagement of participants from the TGD community but did not empirically compare the quality or quantity of engagement with an in-person design process. Future research should examine in-person design sessions compared with web-based design sessions to evaluate and compare the quality and quantity of engagement and interactions. Second, our recruitment method used Facebook to identify participants. This approach yielded a predominantly educated, professional, urban, and white sample. Although we recruited enough TGD individuals for the design of the TGHIR, we acknowledge that this sample may not have been representative of the overall population. As a result, we sought and were awarded funding to test the resources using a diverse research sample. This study is currently underway. Third, the small number of participants was a limitation. In future work, more participants should be involved in the participatory design process to determine whether high levels of web-based participation are maintained in larger samples. Finally, although web-based means of participation helped during the COVID-19 pandemic, a limitation was that it might have also been difficult to reach individuals who did not have the means or ability to access the requisite resources for web-based collaboration, such as high-speed internet and private space [83]. It will be important to keep this accessibility issue in mind when recruiting participants.

Conclusions

Our results had important implications for the use of web-based methodologies in the design of health information resources. Web-based participatory design can support opportunities to contribute despite the potential logistical barriers of in-person design sessions by offering multiple convenient design session times and multiple interaction options. In addition, this approach is helpful when recruiting members from marginalized communities that are small and geographically dispersed, especially rural communities. Not only does the web-based methodological approach work during a pandemic but it may also help when there is historic distrust of research and health care from a community that has been stigmatized and experienced discrimination, such as the TGD community, by researchers and medical providers.

Our evaluation of the web-based participatory design indicated that web-based design sessions can engage participants in creating satisfactory interfaces for accessing and consuming health and medical information. Obtaining web-based input from participants was possible and efficient. Web-based recruitment is also possible for individuals who belong to marginalized communities and provides a platform through which these individuals can safely communicate with others in their community to design health information resources. Integrating web-based platforms can effectively engage participants and yield a positive user experience. Multiple participants reported that a health information resource of this nature would have been helpful in their journey toward gender identity exploration or gender transition.
Acknowledgments
We would like to acknowledge all members and supporters of the gender-diverse community, especially our participatory co-designers and our Community Advisory Board members. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine. This work was supported by funding from the National Institutes of Health, National Library of Medicine, grant GO8 LM013200-01, project name: A Health Information Resource Technology to Reduce Disparities in Transgender Health.

Data Availability
Relevant data associated with, created, and used for the evaluation presented in this paper are openly available from the University of Colorado Strauss Health Sciences Library, Anschutz Medical Campus, Data Repository [84].

Authors' Contributions
BM was responsible for designing and facilitating the participatory design process, coding the qualitative manuscript and web-based interactions, and served as the lead author. AS was responsible for designing the methodology, agile development of the resource, and served as the second author. KY was responsible for obtaining data in usability tests, coding the qualitative transcript and web-based interactions, and reviewing and editing the manuscript. KD was responsible for cataloging the content in the resource, designing the search for credible transgender health resources, and for reviewing and editing the manuscript. MA was responsible for the design of the methodology, the interpretation of data, and for reviewing and editing the manuscript. BDH was responsible for designing the qualitative methodology for focus groups, interpretation of focus group data, and reviewing and editing the manuscript. RSL was responsible for designing the methodology, interpretation of the data, and for reviewing and editing the manuscript. BMK was responsible for designing the methodology, interpretation of the data, and for reviewing and editing the manuscript, and served as the senior author.

Conflicts of Interest
None declared.

References


54. LaMonica HM, Davenport TA, Roberts AE, Hickie IB. Understanding technology preferences and requirements for health information technologies designed to improve and maintain the mental health and well-being of older adults: participatory design study. JMIR Aging 2021 Jan 06;4(1):e21461 [FREE Full text] [doi: 10.2196/21461] [Medline: 3340509]
63. One Colorado - LGBTQ Advocacy Organization. One Colorado. URL: https://one-colorado.org/ [accessed 2021-09-27]
64. Video Conferencing. Cloud Phone, Webinars, Chat, Virtual Events. Zoom. URL: https://zoom.us/ [accessed 2021-06-09]
66. MURAL is a digital workspace for visual collaboration. MURAL. URL: https://www.mural.co/ [accessed 2021-07-09]
67. REDCap. URL: https://www.project-redcap.org/ [accessed 2021-07-09]
Enhancing Exposure Treatment for Youths With Chronic Pain: Co-design and Qualitative Approach

Lea Schemer1, PhD; Courtney W Hess2, PhD; Amanda R Van Orden2, BSc; Kathryn A Birnie3, PhD; Lauren E Harrison2, PhD; Julia A Glombiewski1, PhD; Laura E Simons2, PhD

1Department of Clinical Psychology and Psychotherapy, University of Kaiserslautern-Landau, Landau, Germany
2Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Palo Alto, CA, United States
3Department of Anesthesiology, Perioperative, and Pain Medicine, Department of Community Health Sciences, Alberta Children’s Hospital Research Institute, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Laura E Simons, PhD
Department of Anesthesiology, Perioperative, and Pain Medicine
Stanford University School of Medicine
1070 Arastradero Road
Palo Alto, CA, 94304
United States
Phone: 1 (650) 723 6412
Email: lesimons@stanford.edu

Abstract

Background: Increasing the access to and improving the impact of pain treatments is of utmost importance, especially among youths with chronic pain. The engagement of patients as research partners (in contrast to research participants) provides valuable expertise to collaboratively improve treatment delivery.

Objective: This study looked at a multidisciplinary exposure treatment for youths with chronic pain through the lens of patients and caregivers with the aim to explore and validate treatment change processes, prioritize and develop ideas for improvement, and identify particularly helpful treatment elements.

Methods: Qualitative exit interviews were conducted with patients and caregivers at their discharge from 2 clinical trials (ClinicalTrials.gov NCT01974791 and NCT03699007). Six independent co-design meetings were held with patients and caregivers as research partners to establish a consensus within and between groups. The results were validated in a wrap-up meeting.

Results: Patients and caregivers described that exposure treatment helped them better process pain-related emotions, feel empowered, and improve their relationship with each other. The research partners developed and agreed upon 12 ideas for improvement. Major recommendations include that pain exposure treatment should be disseminated more not only among patients and caregivers but also among primary care providers and the general public to facilitate an early referral for treatment. Exposure treatment should allow flexibility in terms of duration, frequency, and delivery mode. The research partners prioritized 13 helpful treatment elements. Most of the research partners agreed that future exposure treatments should continue to empower patients to choose meaningful exposure activities, break long-term goals into smaller steps, and discuss realistic expectations at discharge.

Conclusions: The results of this study have the potential to contribute to the refinement of pain treatments more broadly. At their core, they suggest that pain treatments should be disseminated more, flexible, and transparent.

(KEYWORDS) co-design; participatory design; pain; exposure treatment; youths with chronic pain; caregivers; qualitative analysis

Introduction

Background

Chronic pain is among the largest contributors to disability in children [1], and suboptimal responses to current treatments remain a challenge for researchers and practitioners [2,3]. Youths with chronic pain experience major barriers to accessing adequate pain treatment (eg, those owing to shortage of providers and geographical distance) [4,5]. When offered a multidisciplinary pain treatment, a substantial number of patients...
In this study, we partnered with patients and caregivers who had previously received GET Living treatment [20]. From an improvement science perspective [15], our aims were to (1) explore and validate treatment change processes, (2) prioritize and develop ideas for improvement (ie, to refine the GET Living program for in-person and remote delivery), and (3) identify particularly helpful treatment elements to promote change.

### Methods

#### Overview and Design

**Overview**

This project comprises two parts: (1) semistructured exit interviews and (2) co-design meetings. Qualitative exit interviews were conducted with patients and caregivers during a discharge session after they received the GET Living intervention as research participants. Subsequent co-design meetings were held with the patients and caregivers as research partners to refine the intervention in a formative research process.

**Setting**

This project involves 2 separate examinations of GET Living: one was a single-arm trial (Boston trial) and the other was an RCT (Stanford trial). The Boston trial (NCT01974791) used a sequential replicated and randomized single-case experimental design (SCED) with multiple measures evaluate the effect of GET on youths with chronic pain for the first time [7]. The Stanford trial (NCT03699007) used a 2-group RCT enhanced with SCED elements to compare GET Living with a traditional multidisciplinary pain management approach [10]. The former GET Living participants had a unique expertise in what it is like to undergo pain exposure treatment from a patient’s and caregiver’s perspective.

#### Recruitment

In the Boston trial, patients were recruited from the Pain Treatment Service at Boston Children’s Hospital between December 2013 and February 2017 (data collection was completed in January 2018). In the Stanford trial, patients were recruited from the Pediatric Pain Management Clinic at Stanford Children’s Health from January 2019 to May 2021 (data collection was completed in January 2022). Treatment providers referred patients to GET Living during clinic visits. A study flyer and additional brochures were also available in the patient waiting room for patients to self-refer to the study. Patients were deemed eligible to participate in GET Living if they were aged 8 to 17 years, had a diagnosis of chronic pain, had moderate to high pain-related fear, and had moderate to high functional disability [7,10].

#### Part 1: Qualitative Exit Interviews

**Goals and Overview**

Interview data were analyzed to identify themes related to treatment change processes. In addition, interviews were conducted to create a pool of ideas for intervention improvement.
and helpful treatment elements, which were later ranked and discussed in the co-design meetings.

**Interviewed Patients and Caregivers**

Only the patients and caregivers who completed all treatment sessions were included in the qualitative analysis to ensure that the data were reflective of the entire treatment experience. The interview that was conducted with a patient and their caregiver who withdrew their participation was excluded. Both the patient and caregiver felt that the treatment’s focus on pain and anxiety was not a good fit. In the Boston trial, 26 interviews of patients and caregivers were analyzed. In the Stanford trial, 26 interviews of patients and caregivers who were randomized to the exposure intervention were analyzed. The patients and caregivers were interviewed separately. More details on the interviewed cohorts are presented in Table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Boston cohort</th>
<th>Stanford cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD; range)</td>
<td>13 (3.12; 8-20)</td>
<td>14 (2.73; 8-18)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>20 (77)</td>
<td>24 (92)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>22 (85)</td>
<td>22 (85)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1 (4)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>2 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Parent marital status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>21 (81)</td>
<td>20 (77)</td>
</tr>
<tr>
<td>Single</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>4 (15)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Widowed</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Pain diagnosis, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>9 (35)</td>
<td>21 (81)</td>
</tr>
<tr>
<td>Neuropathic</td>
<td>8 (31)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Abdominal</td>
<td>6 (23)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Headache</td>
<td>2 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Headache and musculoskeletal</td>
<td>1 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Duration of pain (months), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD; range)</td>
<td>22.6 (27.5; 1-65)</td>
<td>40.5 (37.1; 4-138)</td>
</tr>
<tr>
<td><strong>FDI</strong> at baseline, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD; range)</td>
<td>25.23 (10.3; 2-47)</td>
<td>23.15 (10.07; 4-42)</td>
</tr>
<tr>
<td><strong>Fear of pain (FOPQ&lt;sup&gt;c&lt;/sup&gt; total), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD; range)</td>
<td>50.96 (19.8; 9-82)</td>
<td>56.58 (15.9; 10-84)</td>
</tr>
</tbody>
</table>

<sup>a</sup>GET Living: graded exposure treatment for youths with chronic pain.

<sup>b</sup>FDI: Functional Disability Inventory.

<sup>c</sup>FOPQ: Fear of Pain Questionnaire.

**Interview Guide**

The semistructured exit interviews were conducted by research assistants during the discharge visit following the completion of GET Living. All research assistants were trained by the principal investigator LES. In the Boston trial, most interviews were conducted in person. In the Stanford trial, most interviews were conducted via phone or video calls. The patient and caregiver interview schedules both comprised 8 questions (Multimedia Appendix 1). The questions were intended to capture the positive (eg, question [Q] 1: “What did you like the best about GET Living treatment?” “What was the most helpful?”) and negative (eg, Q2: “What did not help?” “What...
would you change?”) experiences that the families had during their treatment. The participants were also encouraged to give their critical feedback through several questions (eg, Q3: “What do you wish you had known before starting GET Living treatment?”). Other questions targeted to capture treatment change processes, that is, the changes that the families experienced in themselves (eg, Q4: “What did you learn about yourself and your family in GET Living treatment?”). The interview schedule questions guided the conversation; however, consistent with the semistructured nature of the interview, the participants were also provided with space to share additional feedback about their experiences.

Analysis of the Exit Interview Data

Reflexive thematic analysis [21] was used to assess the participants’ perspectives and identify common themes across the interview data. Consistent with constructivist epistemology, reflexive analysis allows for the cocreation of knowledge between the participants and researchers. Subjectivity is not seen as a potential threat to the “truthful” or objective meaning of the data but is rather conceptualized as an analytical resource for data interpretation [22]. Data analysis was led by an investigator (LS) who was not involved in the data collection or intervention delivery. The analysis was conducted by following the 20-question guide by Braun and Clarke [22].

To begin data analysis, the investigator became familiar with the data by repeatedly and actively reading 12 fully transcribed interviews and listening to some randomly selected interviews. For the subsequent coding process, analysis was conducted on the audio recordings of interviews instead of the transcriptions to capture richer, more nuanced (eg, tone and affective aspects of responses) aspects of the participant responses. While listening, the investigator entered detailed notes of the codes for each interview into a comprehensive overview table. Relevant quotes were fully transcribed. Throughout the data analysis, the first author (LS) incorporated semantic features of the data (ie, explicitly stated ideas, concepts, meanings, and experiences) as well as latent features (ie, implicit meanings underlying explicit statements) when defining codes and themes. The generated codes were then clustered into candidate themes. This analytical process focused on the development of themes related to treatment change processes throughout the GET Living program. Theme identification occurred through an iterative process, whereby 2 authors (LS and LES) identified and refined codes and illustrative quotes until deep and nuanced themes regarding change processes were developed. Interview data regarding particularly helpful elements and ideas for improvement were organized into topic summaries (in comparison with fully developed themes). These topics summaries were used as a starting point to facilitate ranking and discussion in the subsequent co-design meetings. They will be presented when describing the results of the co-design meetings.

Part 2: Co-design Meetings

Goals and Overview

The purpose of the co-design meetings was to validate the developed themes related to treatment change processes (eg, regarding their meaningfulness) and reach a consensus regarding important ideas for intervention improvement and key treatment elements. Consensus was established in 6 independent co-design meetings (ie, the nominal group technique) held as 3 parallel meetings with patients and caregivers. This allowed us to establish consensus within groups (ie, consensus in 1 group) and between groups (ie, consensus in multiple groups) as an estimate of the representativeness of the opinions expressed. Patients and caregivers served as ad hoc consultants [15] and were compensated for their efforts (US $30 per hour). Their role was to validate the research findings of the previous thematic analysis and to provide feedback about the GET Living treatment from the receiver’s end [15]. The procedures were preregistered in the Open Science Framework [23]. The GRIPP2 (Guidance for Reporting Involvement of Patients and the Public) checklist for patient and public participation in research guided quality reporting of the study results [24].

Patient and Caregiver Research Partners

Patients and caregivers who were randomized to the exposure treatment arm of the GET Living RCT (Stanford trial), including treatment completers and dropouts, were invited as research partners. Approximately one-third of the people invited accepted the invitation (10/33, 30% patients; 14/33, 42% caregivers). Research partners attended 1 of the 6 independent co-design meetings with parallel meetings for patients (meeting 1a: 4/10, 40%; meeting 2a: 3/10, 30%; meeting 3a: 3/10, 30%) and caregivers (meeting 1b: 4/14, 29%; meeting 2b: 5/14, 36%; meeting 3b: 5/14, 36%). All research partners were invited to a final wrap-up session (5/10, 50% patients and 8/14, 57% caregivers). More details on the research partners who attended the meetings are presented in Table 2.
Table 2. Demographics and pain characteristics of the patient (n=10) and caregiver (n=14) research partners who participated in the co-design meetings.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Youths with chronic pain</th>
<th>Caregivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD; range)</td>
<td>17 (2.4; 10-17)</td>
<td>49 (5.3; 35-55)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10 (100)</td>
<td>12 (86)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>9 (90)</td>
<td>12 (86)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (10)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0 (0)</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>10 (100)</td>
<td>12 (86)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Pain diagnosis, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>8 (80)</td>
<td>N/Aa</td>
</tr>
<tr>
<td>Neuropathic</td>
<td>1 (10)</td>
<td>N/A</td>
</tr>
<tr>
<td>Abdominal</td>
<td>1 (10)</td>
<td>N/A</td>
</tr>
<tr>
<td>Duration of pain (months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD; range)</td>
<td>38.27 (17.3; 14-66)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Procedure

The co-design meetings were scheduled for 2 subsequent calendar weeks (April 2022). The meetings were held via Zoom (Zoom Video Communications, Inc) and lasted approximately 120 minutes (including breaks). An optional web-based wrap-up meeting was held the following week (approximately 60 minutes). An overview of the procedure is provided in Multimedia Appendix 2. The meetings were moderated by CWH, LS, and LES.

Before the meeting, the research partners received a pre-engagement package with an outline of and the materials for the meeting. No preparation was required. At the beginning of the meeting, the research partners introduced themselves with ice-breaking tasks aimed at facilitating a good working atmosphere. Some ground rules were presented. Their role as research partners (as opposed to research participants) was highlighted.

The results of the thematic analysis of treatment change processes were then presented and discussed with the research partners to ensure that the identified themes were relevant and meaningful and to assess whether there were any important change processes missing from the established themes. The ideas for improvement collected during the semistructured exit interviews were then presented. The research partners were asked to rate the ideas using a Qualtrics (Qualtrics International Inc) survey. First, they were asked to select what they believed to be the 10 most important ideas out of the 48 ideas initially identified through the interviews. They were then asked to further refine their initial selection to identify the 3 most important ideas for improvement. The research partners were also encouraged to provide new ideas that were not found on the list as applicable. Once all the answers were collected, the results were shared with the group, and the research partners were asked to discuss the selections to ensure agreement among the retained items and address any differences of opinion regarding key recommendations. The same process was conducted to establish the most helpful out of 38 treatment elements that should be retained in future iterations of the intervention, and where applicable, the research partners were given intervention materials for review.

In the wrap-up meeting, action items from the co-design meetings were presented and finalized in a shared Word (Microsoft Corp) document. The research partners who could not attend the meeting were informed about the action items via email. They were asked to provide their written feedback within 2 weeks.

Evaluation of the Co-design Meetings

At the end of the meetings, all research partners completed module A of the Public and Patient Engagement Evaluation Tool [25]. Module A was developed to measure a 1-time engagement activity from a participant’s perspective. The module consists of 13 statements (eg, “I had a clear understanding of the purpose of the co-design meeting”), which the research partners were instructed to rate on a 5-point Likert scale (1=“strongly disagree,” 2=“disagree,” 3=“neither agree nor disagree,” 4=“agree,” and 5=“strongly agree”). The questionnaire also comprises 6 open-ended questions addressing
key elements of quality public and participant engagement, including the integrity of design and process, influence and impact, participatory culture and collaboration, and common purpose. The questionnaire was used as a quality measurement of research partner engagement in the co-design meetings. In addition, we openly asked the research partners why they agreed to participate in the co-design meetings.

**Ethics Approval**

Both trials received ethics approval from their respective institutional review boards (Boston: IRB-P0000727, and Stanford University: Protocol 39514). Before their participation, the patients and caregivers actively consented to take part in the respective clinical trial. The final version of the manuscript was sent to the patient and caregiver research partners. All research partners provided their consent for publication.

**Results**

Part 1: Results of the Thematic Analysis

A total of 3 subordinate themes were generated from the reflexive thematic analysis of the exit interviews (Figure 1). These themes reflect the treatment change processes experienced by the patients and caregivers during the GET Living intervention. The themes were validated by the research partners in the co-design meetings.

**Figure 1.** Subordinate themes describing treatment change processes experienced by the patient and caregiver. The developed themes are summarized on the left. The subthemes are displayed within the boxes, with the subthemes derived for the patients presented on the left (“I didn’t think I would be able to jog and I was able to do it” and “I am actually capable”), the subthemes derived for the caregivers presented on the right (“Chronic pain is overwhelming,” “Learn to relinquish the control,” and “I understand in more detail”), and the subthemes derived both patients and caregivers presented in the middle (“Broadening the toolkit” and “I have a support system”). GET: graded exposure treatment.

**Theme 1: GET Living Helps Process Pain-Related Emotions**

**Overview**

The first theme described how the patients and caregivers were better able to handle their pain-related emotions. Although the patients felt more confident in dealing with challenging situations, the caregivers had a space to process their own emotional struggles.

**Patients: I Didn’t Think I Would Be Able to Jog and I Was Able to Do It**

The patients learned through exposures that the experience was not as bad or challenging as they thought it would be. Overall, the patients described an emotional shift in their experience because it did not match their expected outcome:

> It was probably the first time they told me to go for a jog, I didn’t think I would be able to do it, I got really scared but after I jogged with my mum it made me feel a lot better and I was able to do it and it made me happy. [B9, patient]

Other patients and caregivers reported a change in their thinking:

Exposures were easier than I thought they would be. [B30, patient]

The thought process going into it and getting those thoughts in check of how you can do things and not allowing misguided thinking to not allow you to do things that you can do. [S24, caregiver]

There was a general shift in their perception of challenges. The patients also appeared to gain a sense of control:

> I learned fear doesn’t control me, I can control it and I can control how to deal with it. [S15, patient]

I learned to not be so afraid of things I loved to do. There are some challenges but I can get through them. [B20, patient]

Those exposures really assisted her achieving goals she didn’t think she’d be able to make. And after that she was able to do more things. [S24, caregiver]

Taken together, the patients realized that the situations they once feared were not as emotionally challenging or difficult as they expected them to be. On the basis of this experience, the patients seemed to see challenges as more approachable and manageable:
**Caregivers: Chronic Pain Is Overwhelming**

The caregivers became more aware of the overall experience of managing chronic pain and its impacts on the entire family:

> Chronic pain is so overwhelming and such a challenge. Not only for the person in pain but the entire family really suffers from that...I don’t think I recognized how bad it was until I got into the program. [B19, caregiver]

> I learned that my pain impacted everyone in my family not just me, so like, the pain, I might feel it but everyone else can experience it too. [B29, patient]

The caregivers had room to express and process their own emotional struggles:

> And some were just the sad or grieving things. Especially her age going into young adulthood. [S46, caregiver]

> And [clinicians] said, “Your daughter’s gonna be okay. Her pain is real, but she’s gonna be ok.” That was really important for me to hear that. Finally, someone put it all together and made me feel like, okay she’s not gonna break. [GS33, caregiver]

Some even expressed a newly found admiration toward their children:

> So, I guess I admire kids and people who get through the pain somehow, and it’s without a break and they still manage. I guess my admiration for [child] and for people who’ve experienced that has increased. [B33, caregiver]

> And then, you know, I knew that [child] had it in him that he could push himself. I just think he needed help kinda pushing himself past that initial pain. [S9, caregiver]

Taken together, the overwhelming experience of living with chronic pain was felt by the entire family. The caregivers mostly expressed feeling anxious or sad for their child; however, they were able to shift their perception by reinterpreting this struggle as a strength. Instead of feeling sad or anxious, they expressed an admiration for their children for handling painful situations.

**Theme 2: GET Living Empowers Patients and Caregivers**

**Overview**

The second theme described how the patients and caregivers felt empowered during treatment. Whereas the patients experienced becoming more confident, the caregivers gave their children more space to handle difficult situations by themselves. Both felt that they learned concrete strategies for navigating difficult situations.

**Patients: I Am Actually Capable**

The patients’ experiences changed their perceptions of themselves. Many patients felt empowered and more confident. Some patients learned that they were capable of doing things despite their pain:

> The only thing stopping me was myself...Well of course it was my back pain and all that. But I kind of held on to my back pain a little too much, for a little bit too long. [S33, patient]

> I learned that because I’ve been in a lot of pain and I put things off, that I am actually capable of doing a lot more. [B25, patient]

> I also learned that I can do anything even with my neck pain. [S24, patient]

Other patients expressed a more generalized sense of being capable:

> I am stronger than I thought I was. [B31, patient]

> I got more confident and stronger doing all the activities. [B22, patient]

> I learned that I was more determined and stronger than I thought. [S17, patient]

Taken together, the patients changed their self-perception and appeared to be more confident in their ability to handle difficult situations in the face of pain and other challenges.

**Caregivers: Learn to Relinquish the Control**

The caregivers came to understand how to balance control and letting go. Some caregivers expressed that they could better see the benefits of giving their children more opportunities to handle their pain by themselves:

> As parents, we do want to help out and control as much as we can, and to some extent I do still believe that we should be looking out for each other, you know, trying to prevent them from having pain, if it’s possible. But if it’s, in this kind of situation with the chronic pain thing, you learn to relinquish the control more and give them more options to handle it themselves. [S3, caregiver]

> I learned that [child] can be a lot more independent than I sometimes give him credit for. So sometimes I have to ease off in helping him. So, I think I learned that it is okay to let him tumble through something because then he will feel like he really did it himself. [S60, caregiver]

Other caregivers reported that they became more aware of the negative effects of being overly controlling:

> Well obviously, that we were holding her back from trying new things and not presenting things that would challenge or take her outside of her comfort zone. And we didn’t realize what we were doing. [B19, caregiver]

> I learned that I can be pretty intense and anxious which contributes to my child’s troubles or doesn’t help her cope. I learned to be more relaxed. Wasn’t too rigid before, let go of that now. [B5, caregiver]

Generally, the caregivers were able to hold back on responding with their initial reaction to better respond to the specific needs of their children:

> For me it was really about my own responses to her and how to control my responses and be more...
The caregivers felt that they had solid action plans to encourage long-term goals:

The patients learned to break down activities in pursuit of concrete strategies to better cope and live with chronic pain:

But also offering additional strategies for feeling like we don’t have to be helpless in the face of the pain when it’s severe. I think that was super helpful to both of us. [S22, caregiver]

I liked the emphasis on creating a sense of confidence and broadening the toolkit for dealing with pain and the headaches. Kind of finding ways to carry forward. To make life worth living, kind of; you know, preserve some quality of life. [S22, caregiver]

With strategies of knowing how to cope with pain, okay, I am gonna be able to approach those challenges of life more so because I have that tool under my belt that says, oh you are hurting today, you are not having a good day, how are you gonna get going and in the end come out the other side. [B23, caregiver]

The patients learned to break down activities in pursuit of long-term goals:

I learned that taking a [small] step at a time can help me improve so much more than trying to take a big step. [B36, patient]

I learned to just make accommodations instead of stopping the activity altogether. [B30, patient]

I think her learning how to set goals that are achievable and measurable. And for her to be able to make them so that they are realistic. So, it was really her individualized goals. She made them up and she decided with the team where her values were. [B19, caregiver]

The caregivers felt that they had solid action plans to encourage activities:

To give me a bit more vocabulary or instructions how to talk about things, like... “remember you wanted to do this because of the goals you set for yourself.” [S60, caregiver]

And maybe just validate that we get it and when she does something that we recognize that. It does really help because when you are in the middle of it, you don’t really think about how we are going to react and it changes how she feels. [B19, caregiver]

It empowered us as parents to say, we know your pain is real, we know it might cause a little bit of back pain, and you can take breaks. It gave us strategies for what we can say and what we can do to help encourage her still do her everyday activities. [S33, caregiver]

Taken together, the patients and caregivers learned concrete strategies to navigate through difficult situations. These strategies helped reduce feelings of helplessness in both the patients and caregivers.

**Theme 3: GET Living Improves Family Relations**

Overview

The third theme described how GET Living helped improve the relationship between patients and their caregivers. Being able to better understand the complexity of chronic pain, the caregivers were more able to validate their child’s experiences and felt closer to them. The patients also indicated that they felt more supported by their caregivers.

Caregivers: I Understand in More Detail

The caregivers better understood their child’s pain experience in their day-to-day difficulties:

I knew that she was hurting every day and that lots of things were difficult for her, but I think that I understand in more detail that even simple tasks, how and why they are difficult for her. [S57, caregiver]

[Clinicians] taught me a lot regarding just [child]’s pain and how it can really, I don’t know, change her behavior. In that if [child] is grumpy or tired. I never associated the pain with her emotions before, neither did my husband. So, it was really eye opening for us to understand the correlation. [S24, caregiver]

The caregivers also understood the driving mechanisms of pain chronicity in more detail:

I was also kind of surprised in the session when he was doing the soccer practice because he kind of attributed the time when all his leg pain started with soccer, even though soccer did not, you know, cause it. [S60, caregiver]

It took away my anxiety that it will hurt, but it won’t harm her. The program made her try something. And some of the things she did, I knew that she would hurt herself. Not harm herself, but hurt herself. [B19, caregiver]

The caregivers also reported being better able to validate the experiences of their child:

It never dawned on me before about how [child] could be feeling about this because no one can see it. And we just gave her a hard time about school and that she is not feeling it. And sometimes with that they have to keep validating it and hold on to it. And maybe just validate that we get it and when she does something that we recognize that. It does really help because when you are in the middle of it, you don’t...
really think about how we are going to react and it changes how she feels. [B19, caregiver]

My mom and dad...actually knew how I feel now and what I was going through. [B20, patient]

Taken together, the caregivers became more aware of the difficulties of their children. They better understood the impact of chronic pain on pain-related disability and distress. The caregivers also became aware of the driving mechanisms of chronic pain, including emotional responses and misattributions. This allowed them to validate their child’s experience more.

Patients and Caregivers: I Have a Support System

The patients and caregivers reported that GET Living fostered improved family connections. The patients became more aware that they were not alone because they had their caregivers and families to support them:

And that my family can help me do whatever, that I don’t just have to rely on myself to help these things. I have a support system. [S33, patient]

I learned that my family are very enthusiastic and willing to do those things with me. [S22, patient]

And my family, I think, learned if I am in pain how they can help me deal with it. [S15, patient]

The caregivers also felt a closer connection with their children:

We kind of had a better connection than we did before. Not that we had a bad connection, it’s just the drives to the sessions. [S9, caregiver]

I felt like some of the sessions led to more discussions with [child] and I afterwards, like I felt that there were certain things, like as a mother daughter, that it was positive. [S46, caregiver]

Taken together, the relationships between the patients and their families improved. While the patients felt supported, the parents felt a closer connection with their children.

Part 2: Results of the Co-design Meetings

Ideas for Improvement

A total of 12 ideas for improvement were prioritized in multiple groups (ie, between-group consensus) and are presented in Table 3. The ideas were organized based on the degree of consensus between the groups. Five ideas that were prioritized by within-group consensus are presented in Multimedia Appendix 3.

Interestingly, these improvement ideas were not specific to exposure treatments and could be applied to any form of behavioral or physical pain treatment. For example, the research partners agreed that pain exposure treatment should be disseminated more. There was absolute consensus (consensus in 6/6, 100% co-design meetings) that pain exposure treatment should use patient testimonials to (1) provide patients with narratives of how other patients are dealing with similar difficulties, (2) inform future patients about what treatment will be like, (3) provide a role model, and (4) promote positive expectations. Most research partners also agreed that more efforts should be made to create awareness among the general public and primary care providers to facilitate an early referral for treatment (consensus in 4/6, 67% co-design meetings).
<table>
<thead>
<tr>
<th>An ideal GET Living&lt;sup&gt;b&lt;/sup&gt; program would...</th>
<th>Concrete ideas</th>
<th>Consensus between groups (n=6), n (%)</th>
</tr>
</thead>
</table>
| ...inform what the treatment will be like and promote positive expectations | • Patient and caregiver testimonials (eg, videos) to see other patients dealing with similar difficulties, provide a role model, better understand what the treatment will be like, transmit hope for future patients  
  • Clarify that the treatment aims to increase activity and explain the role of PT<sup>c</sup> (compared with traditional PT) | 6 (100) |
| ...start earlier with more interdisciplinary exchange | • More awareness of the program through posters, flyers, websites, and social media  
  • Campaign educating primary care providers about this modality as a treatment option to facilitate early referral  
  • More exchange and referral between providers (eg, to discuss treatment progress) | 4 (67) |
| ...allow for more flexibility | • Adapt the duration, frequency, and content to the momentary pain level or energy of patients  
  • Flexible web-based sessions when pain level is too high | 4 (67) |
| ...be also offered remotely with optional in-person meetings | • Optional in-person meetings to build trust and help patients get a better diagnostic view of the exposure activities  
  • Help overcome technical barriers (eg, send treatment materials at home and provide Wi-Fi booster) | 3 (50) |
| ...add booster sessions | • Combination of structured and client-lead booster sessions (eg, reminder of the core treatment elements and think together how they can be applied to real life) | 3 (50) |
| ...be honest that becoming better is not easy but it is a process | • Emphasize that treatment provides long-term strategies  
  • Provide feedback on progress (especially little steps) as a motivator  
  • Help to find the balance of being challenged but not overwhelmed | 3 (50) |
| ...have the patient decide if parent should participate in treatment | • Discuss with patients whether caregivers should join the treatment  
  • Optional patient-only sessions | 3 (50) |
| ...be offered also to patients over 18 | • Support in an especially vulnerable phase of transition into young adulthood (eg, decision on the future) on top of chronic pain | 2 (33) |
| ...enable patients to meet other patients | • Platform to exchange information with other patients of similar age (eg, ages of 8 to 12 years and ages 13 to 17 years)  
  • Open coffee hours via Zoom (eg, once per month)  
  • Web-based education sessions or booster sessions together with other patients | 2 (33) |
| ...enable parents to meet other parents | • Platform for support and exchange (eg, see other families who go through the same thing and think together how to positively influence family dynamics) | 2 (33) |
| ...include more complex pain ratings | • Description of end points (eg, developing individualized reference points at the beginning of treatment)  
  • Body map to describe pain localization and give differential pain ratings for different locations | 2 (33) |
| ...be adapted to other symptoms experienced besides pain | • For example, adapt exposure activities to additional symptoms of dizziness  
  • Editable worksheets to personalize exposure activities | 2 (33) |

<sup>a</sup>Ideas for improvement are organized according to the degree of consensus between groups.

<sup>b</sup>GET Living: graded exposure treatment for youth with chronic pain.

<sup>c</sup>PT: physical therapy.

**Helpful Treatment Elements to Promote Change**

A total of 13 treatment elements were considered helpful in promoting change in multiple groups (between-group consensus; Table 4). For a clear overview, helpful treatment elements are organized by treatment phase. Seven treatment elements that were considered helpful by only the members of 1 group are presented in Multimedia Appendix 4 (within-group consensus).

In general, the research partners appreciated the understanding attitude of clinicians, personalization of treatment through the...
pursuit of individualized goals, education about chronic pain, encouragement of activities, and discussion of realistic expectations at discharge. For example, during the phase of goal setting, a majority of the research partners agreed that future exposure treatments should continue to empower patients to be “in charge” to choose meaningful exposure activities (consensus in 5/6, 83% co-design meetings), break long-term goals into smaller steps (consensus in 5 of the 6 co-design meetings, 83%), and help patients become aware of their own values and motivators (consensus in 2/6, 33% co-design meetings).

Table 4. Most helpful treatment elements agreed upon in co-design meetings using the nominal group technique (consensus in multiple groups).

<table>
<thead>
<tr>
<th>Treatment phase and future GET Living programs (regardless of the delivery format) should continue to...</th>
<th>Consensus between groups (n=6), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Building rapport</strong></td>
<td></td>
</tr>
<tr>
<td>...combine pain psychology and physical therapy</td>
<td>3 (50)</td>
</tr>
<tr>
<td>...transmit the feeling that it is possible to deal with pain</td>
<td>2 (33)</td>
</tr>
<tr>
<td>...offer validation and understanding of patients’ situation</td>
<td>2 (33)</td>
</tr>
<tr>
<td><strong>Goal setting</strong></td>
<td></td>
</tr>
<tr>
<td>...empower patients to be “in charge” to choose meaningful activities</td>
<td>5 (83)</td>
</tr>
<tr>
<td>...distinguish between short-term and long-term goals</td>
<td>5 (83)</td>
</tr>
<tr>
<td>...help patients become aware of their values and motivators</td>
<td>2 (33)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>...reflect on triggers of pain and anxiety</td>
<td>3 (50)</td>
</tr>
<tr>
<td>...distinguish between short-term and long-term solutions</td>
<td>2 (33)</td>
</tr>
<tr>
<td>...include the exposure graphs</td>
<td>2 (33)</td>
</tr>
<tr>
<td><strong>Exposures</strong></td>
<td></td>
</tr>
<tr>
<td>...encourage activities allowing for breaks and a slow pace</td>
<td>3 (50)</td>
</tr>
<tr>
<td>...teach the use of facilitators</td>
<td>3 (50)</td>
</tr>
<tr>
<td>...include the WILD scale</td>
<td>2 (33)</td>
</tr>
<tr>
<td><strong>Discharge</strong></td>
<td></td>
</tr>
<tr>
<td>...discuss realistic expectations at discharge (eg, discuss coping with pain flare-ups)</td>
<td>6 (100)</td>
</tr>
</tbody>
</table>

*Treatment elements that were considered helpful are organized by treatment phases.

GET Living: graded exposure treatment for youth with chronic pain.

The WILD scale assesses a patients’ perceived Willingness, Importance, Likelihood of Success, and Difficulty with regard to the chosen exposure. The scale is completed before and after exposure [10]. The WILD scale of an example patient can be found in Multimedia Appendix 5.

**Evaluation of the Co-design Meetings**

Overall, the co-design meetings were evaluated as good, with mean values being consistently at the upper end of the agreement scale (Tables 5 and 6). The research partners felt that the co-design meeting was a good use of their time, that they were able to contribute, and that they were confident that the meeting’s goals were achieved.
Table 5. Quantitative results of the Public and Patient Engagement Evaluation Tool\(^a\).

<table>
<thead>
<tr>
<th>Item</th>
<th>Patients, mean (SD; range)</th>
<th>Caregivers, mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Communications and supports for participation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had a clear understanding of the purpose of the co-design meeting.</td>
<td>4.3 (0.95; 2-5)</td>
<td>4.4 (0.63; 3-5)</td>
</tr>
<tr>
<td>The supports I needed to participate were available (eg, travel, childcare, etc.).</td>
<td>4.5 (0.71; 3-5)</td>
<td>4.4 (0.76; 3-5)</td>
</tr>
<tr>
<td>I had enough information to contribute to the topic being discussed.</td>
<td>4.6 (0.52; 4-5)</td>
<td>4.6 (0.51; 4-5)</td>
</tr>
<tr>
<td><strong>Views and perspectives</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was able to express my views freely.</td>
<td>5 (0; 0-5)</td>
<td>4.9 (0.36; 4-5)</td>
</tr>
<tr>
<td>I feel that my views were heard.</td>
<td>4.9 (0.32; 4-5)</td>
<td>4.9 (0.36; 4-5)</td>
</tr>
<tr>
<td>A wide range of views on the topics discussed was shared.</td>
<td>4.5 (0.53; 4-5)</td>
<td>4.7 (0.47; 4-5)</td>
</tr>
<tr>
<td>The individuals participating in this co-design meeting represented a broad range of perspectives on the topic.</td>
<td>4.4 (0.7; 3-5)</td>
<td>4.5 (0.52; 4-5)</td>
</tr>
<tr>
<td><strong>Impacts and influence of engagement initiative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that the co-design meeting achieved its objectives.</td>
<td>4.7 (0.48; 4-5)</td>
<td>4.5 (0.52; 4-5)</td>
</tr>
<tr>
<td>I am confident the input provided through this initiative will be used by Biobehavioral Pediatric Pain Lab.</td>
<td>4.5 (0.71; 3-5)</td>
<td>4.6 (0.65; 3-4)</td>
</tr>
<tr>
<td>I think the input provided through this activity will make a difference to the work of the Biobehavioral Pediatric Pain Lab.</td>
<td>4.6 (0.7; 3-5)</td>
<td>4.6 (0.5; 4-5)</td>
</tr>
<tr>
<td><strong>Final thoughts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As a result of my participation in the co-design meeting, I am better informed about the Biobehavioral Pediatric Pain Lab.</td>
<td>4.4 (0.7; 3-5)</td>
<td>4.1 (0.77; 3-5)</td>
</tr>
<tr>
<td>Overall, I was satisfied with this engagement initiative.</td>
<td>4.6 (0.52; 4-5)</td>
<td>4.7 (0.47; 4-5)</td>
</tr>
<tr>
<td>This engagement initiative was a good use of my time.</td>
<td>4.7 (0.48; 4-5)</td>
<td>4.6 (0.65; 3-5)</td>
</tr>
</tbody>
</table>

\(^a\)1=strongly disagree, 2=disagree, 3=neither disagree or agree, 4=agree, and 5=strongly agree.
### Table 6. Qualitative results of the Public and Patient Engagement Evaluation Tool plus reasons for participation.

<table>
<thead>
<tr>
<th>Open-ended questions</th>
<th>Patients</th>
<th>Caregivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>What else would you like us to know about how your participation in the co-design meeting was supported?</td>
<td>• Felt supported</td>
<td>• Easy web-based format with handouts given before</td>
</tr>
<tr>
<td>What else would you like us to know about how you were able to share your views?</td>
<td>• Everyone was easy to talk to</td>
<td>• Accommodating and flexible scheduling</td>
</tr>
<tr>
<td></td>
<td>• Everyone brought different perspectives and life experiences, which shaped their advice and made the discussion interesting</td>
<td>• Easier to share openly via Zoom</td>
</tr>
<tr>
<td>What else would you like us to know about the influence you think the co-design meeting will have?</td>
<td>N/A</td>
<td>• Input may guide future improvements of an already great program</td>
</tr>
<tr>
<td>What were the strengths of the co-design meeting?</td>
<td>• Everyone was nice and supportive</td>
<td>• Leaders were open and understanding, and our opinions were validated</td>
</tr>
<tr>
<td></td>
<td>• Ability to contribute perspective on what to improve upon</td>
<td>• Valuable to hear other perspectives from other patients</td>
</tr>
<tr>
<td></td>
<td>• Engaging and friendly leaders</td>
<td>• Able to voice concerns and connect with and hear the opinions of other caregivers</td>
</tr>
<tr>
<td></td>
<td>• Materials provided in advance</td>
<td>• Smaller groups allowed for everyone’s voice to be heard</td>
</tr>
<tr>
<td></td>
<td>• Surveys helped facilitate discussion</td>
<td>• Breakout rooms so that youths and caregivers could discuss separately</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Informal nature allowed for comfortability</td>
</tr>
<tr>
<td>What could be improved about the co-design meeting?</td>
<td>• More icebreakers and introductions to meet the others in the meeting</td>
<td>• Would have liked a time to share freely without any structure</td>
</tr>
<tr>
<td></td>
<td>• Allow the patients to talk freely about their experience without structure to allow for suggestions that the researchers had not proposed and to allow the patients to connect with one another</td>
<td></td>
</tr>
<tr>
<td>What else would you like us to know about your experience with the co-design meeting?</td>
<td>• It was a great way to allow the past patients to feel more included and important</td>
<td>• Allowed caregivers to hear others’ experiences and thoughts</td>
</tr>
<tr>
<td>Why did you agree to be part of the co-design meetings?</td>
<td>• Wanted voice to be heard</td>
<td>• To help others with chronic pain</td>
</tr>
<tr>
<td></td>
<td>• Wanted to give back to a program that helped me</td>
<td>• To give back and help this program</td>
</tr>
<tr>
<td></td>
<td>• Wanted to help improve the program for others with chronic pain</td>
<td>• This study was very important to our family</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Wanted to share my ideas for improvements</td>
</tr>
</tbody>
</table>

No one answered the question.

**Discussion**

**Principal Findings**

**Overview**

This study looked at a multidisciplinary exposure treatment for youths with chronic pain through the lens of patients and caregivers. First, qualitative analysis of exit interviews conducted with patients and caregivers after they received the GET Living intervention explored the treatment change processes. Second, co-design meetings with patients and caregivers as research partners aimed to refine the GET Living intervention. The implications of both aspects are discussed in the subsequent sections.

**Treatment Change Processes: What Changes and How?**

The qualitative analysis revealed a wide range of treatment change processes, indicating that what happens within patients during treatment is complex and difficult to describe from a single theoretical lens [19]. The patients and caregivers described that the exposure treatment helped them to (1) better process pain-related emotions, (2) feel empowered, and (3) improve their relationship with each other. The elements of these reported changes align with different theoretical models. In line with the inhibitory learning approach [26,27], the patients experienced a violation of their expectations, wherein feared situations were not as emotionally challenging or difficult as expected they them to be. By contrast, the caregivers reported a reduction in protective behavioral responses when they felt more in control of their emotional distress, which, in turn, empowered the patients to handle difficult situations themselves. This aligns with the theoretical assertions of the interpersonal...
The research partners conveyed that increasing the access to pain treatments more broadly. In terms of their family and social environment, they felt more supported (“I have a support system”). Consistent with the interpersonal process model of intimacy, the patients and caregivers experienced an increase in intimacy and improvement in their relationship when the caregivers were better able to understand and validate the patients’ pain experience [31]. Looking through the lens of self-determination theory [32], the treatment might have satisfied the need for autonomy and competence (eg, by having patients be “in charge,” which was ranked as a particularly helpful treatment element), which facilitated goal pursuit despite chronic pain and an increased a sense of confidence. The patients also felt more supported, indicating satisfaction in the need for relatedness. However, the patients wished to extend this support to their peers with chronic pain. Altogether, the present results underscore the need for a more holistic approach to understand the full complexity of treatment change processes within patients and in their interaction with their social environment. Future research using and combining contemporary quantitative methods (eg, ambulatory assessments, network analyses, and SCEDs) could use the present findings to flexibly and rigorously study treatment change processes from idiographic and nomothetic perspectives [33].

**Refinement of Pediatric Pain Treatments: What Should We Do Better?**

The research partners prioritized 13 core treatment elements that were helpful in promoting change. This feedback can directly inform clinicians which specific behaviors and techniques are perceived as impactful. This is informative for clinicians in general but especially in settings with time constraints. For example, a majority of the research partners agreed that future exposure treatments should continue to assist patients in finding and pursuing meaningful goals. This recommendation agreed with pain scientists, who advise combining exposure treatment with clarification interventions to identify personal goals and goal conflicts [34]. Such techniques could also have the potential to ameliorate other behavioral programs. However, future research should systematically investigate the benefits of these techniques (eg, improving outcomes or facilitating the transfer of skills to daily life).

The research partners also agreed upon 12 ideas for improvement. At their core, these ideas suggest that pain treatments should be disseminated more, flexible, and transparent. The research partners advised that there should be a platform for exchange between people with lived experiences and that the complexity of the individual pain experience should be acknowledged. To our surprise, most ideas were not specific to the content and refinement of exposure treatment; instead, they could inform the implementation of behavioral or physical pain treatments more broadly.

The research partners conveyed that increasing the access to and dissemination of pain psychology treatment is of upmost importance, a message also building momentum among pain scientists [35]. At the receiver’s end, the research partners recommended to better clarify the role of psychological interventions in the context of a multidisciplinary pain treatment approach. They, therefore, came up with creative ideas such as video testimonials or advertising campaigns to clarify treatment aims and promote positive expectations. At the same time, the research partners also suggested better acquainting other treatment providers with this treatment option to facilitate early referral. In addition, they considered an increase in flexibility (eg, in terms of session duration, frequency, content, and delivery format depending on momentary pain level) a promising step toward improvement. Although shifting plans based on pain levels stands in contrast to pain scientists advocating that time and quota–contingent treatment plans are preferred over pain-contingent plans [36], it introduces an important consideration for pragmatic implementation in real life. Momentarily scaling back an activity versus rigid adherence to a plan could ultimately provide the flexibility needed to reach the long-term goal of greater life engagement and functionality.

The research partners considered the remote delivery format with optional in-person check-ins (eg, to build trust) as promising beyond the pandemic, which aligns with initiatives underway in the pain treatment field, as the pandemic has accelerated the dissemination of remotely delivered pain management services [37]. Continuing this path might contribute to a greater dissemination of pain psychology treatments, especially among youths. Remotely delivered treatments might also be beneficial for other behavioral or physical treatments (eg, to facilitate integration into daily life). Moreover, the research partners wanted more support in transferring and maintaining learned strategies (eg, via booster sessions). This request suggests potential ways to address the issue that the effects of pain psychology treatments are often not stable over time [3]. Altogether, the research partners created an abundant set of ideas focused on improving the delivery of pain treatments. From a human-centered design perspective, the present results specify the needs of patients and caregivers [38]. Future research could use these ideas to investigate whether tailoring implementation strategies to end users’ needs relates to better behavioral (eg, penetration) and perceptual (eg, acceptability) implementation outcomes [39]. For example, it would be interesting to see whether tailoring implementation strategies for pain treatments results in fewer people declining to participate and fewer dropouts.

**Strength and Limitations**

We provided an in-depth analysis of a specialized multidisciplinary exposure treatment for youths with chronic pain. Although we provided an overview of the change processes experienced by patients and caregivers, we could establish whether they contributed to the overall improvement (eg, increase in physical activity and school performance) using the methods we adopted. The patients and caregivers did not report having experienced treatment side effects, although it should be noted that we included only treatment completers in the thematic analysis. We also did not explicitly ask about treatment side effects. Our findings may not be generalizable to other behavioral pain treatments or pain populations, although it is...
likely that the fundamental processes identified are cross-cutting. Multiple co-design meetings allowed us to establish consensus within and between groups. This can be taken as an estimate of the representativeness of the expressed opinions. However, the included research partners were not representative, even of the US population, in terms of underrepresented groups, with most research partners being White and female. The study was conducted within the US health care system, and the results may not be generalizable to other health care systems and countries. In Germany, for example, the distances between patients’ homes and outpatient pediatric care centers are smaller, and the acceptance of internet-delivered treatments is rather low [40]. We did not present differential consensus ratings for patient and caregiver research partners because they were largely congruent. Only the involvement of caregivers during treatment was a critical point, where although the patients wanted less involvement, their parents wanted more involvement. The compromise developed in the wrap-up meeting was to negotiate the amount of involvement at the beginning of treatment (also depending on the patient’s age) and offer patient-only sessions.

The Future of GET Living

The GET Living team is poised to iterate and implement the advice learned from the patients and caregivers as research partners. Planned modifications span 3 key domains: publicity and education, treatment delivery, and supporting families after treatment completion. We intend to develop video testimonials that weave in the ingredients the patients and caregivers defined as essential, namely the opportunity to process pain-related emotions, feeling empowered, and improving their relationship with one another. In addition to patient and parent testimonials, we would like to roll out an advertising campaign that targets both patient families and providers regarding the role of psychology in pain treatment and in some instances, more specifically, the GET Living treatment approach. These 2 publicity and education initiatives will better elucidate treatment aims, address misconceptions, and cultivate positive expectations regarding treatment. In the realm of treatment delivery, we have demonstrated in our latest clinical trial the capability to deliver GET Living remotely [11], and a clinical trial to implement a digital exposure intervention is underway ([41]; NCT05079984). Finally, we aim to devise approaches that will lead to lasting positive effects. We envision integrating booster sessions up to 1 year after treatment completion, potentially a combination of in-person and remotely delivered sessions. Moreover, we can leverage our developing digital content to push resources to patient families over time and provide a library of tools accessible long after treatment completion. Altogether, these research partner–guided changes will undoubtedly improve engagement and outcomes among youths with chronic pain. For the future of GET Living, we plan to establish patients and caregivers as standing members of an advisory board to facilitate a closer collaboration with them during the next iteration of GET Living.

Conclusions

This study has revealed several powerful implications that should be considered in future treatments and studies. The exit interviews with the patients and caregivers demonstrated the full complexity of treatment change processes. The research partners agreed that pain exposure treatment should be disseminated more, flexible, and transparent. These implications would not have been revealed if only traditional outcome and facility measures had been used. The clear and meaningful outcomes of this study strongly support the involvement of patients and caregivers in pain treatment manual developments and pain study designs.

Acknowledgments

The Boston trial was supported by the American Pain Society and Sharon S Keller Chronic Pain Research Grant and Deborah Munroe Noonan Memorial Research Fund; the Stanford trial was supported by NIAMS/R21AR072921; and the co-design meetings were supported by NIAMS/K24AR078945 awarded to LES. The authors thank Alex Tremblay-McGaw, Ellison Choate, and Nicole Jehl for their help during the co-design meetings.

Authors’ Contributions

LS and LES developed the ideas for the manuscript. LS conducted the thematic analysis, supervised by LES. LS and LES designed and planned the co-design meetings together with CWH, ARVO, KAB, and LEH. LS, LES, CWH, and ARVO conducted the co-design meetings. LS drafted the first version of the manuscript with inputs from all the authors. CWH mainly revised aspects of the thematic analysis. KAB mainly revised aspects of the patient engagement process. LEH and JAG mainly provided input on clinical implications. LES supervised the project. All the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of semistructured interview questions.

[PDF File (Adobe PDF File), 104 KB - jopm_v15i1e41292_app1.pdf ]

Multimedia Appendix 2
Co-design meeting schedule.

Multimedia Appendix 3
Ideas of improvement developed and agreed upon in co-design meetings using the nominal group technique (consensus in one group).

Multimedia Appendix 4
Most helpful treatment elements agreed upon in co-design meetings using the nominal group technique (consensus in one group).

Multimedia Appendix 5
The Willingness, Importance, Likelihood of Success, and Difficulty scale.

References


Abbreviations

- GET: graded exposure treatment
- GRIPP2: Guidance for Reporting Involvement of Patients and the Public
- RCT: randomized controlled trial
- SCED: single-case experimental design
- Q: question

© Lea Schemer, Courtney W Hess, Amanda R Van Orden, Kathryn A Birnie, Lauren E Harrison, Julia A Glombiewski, Laura E Simons. Originally published in Journal of Participatory Medicine (https://jopm.jmir.org), 09.03.2023. 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 Journal of Participatory Medicine, is properly cited. The complete bibliographic information, a link to the original publication on https://jopm.jmir.org, as well as this copyright and license information must be included.
Acceptability of Automated Robotic Clinical Breast Examination: Survey Study

George P Jenkinson¹*, MSc; Natasha Houghton²*, MBChB; Nejra van Zalk³, PhD; Jo Waller⁴, PhD; Fernando Bello⁵, PhD; Antonia Tzemanaki¹, PhD

¹Bristol Robotics Laboratory, Department of Mechanical Engineering, University of Bristol, Bristol, United Kingdom
²Centre for Engagement and Simulation Science, Department of Surgery and Cancer, Imperial College London, London, United Kingdom
³Dyson School of Design Engineering, Imperial College London, London, United Kingdom
⁴Cancer Prevention Group, School of Cancer & Pharmaceutical Sciences, King’s College London, London, United Kingdom
*these authors contributed equally

Corresponding Author:
George P Jenkinson, MSc
Bristol Robotics Laboratory
Department of Mechanical Engineering
University of Bristol
T Block, UWE Frenchay Campus
Coldharbour Lane
Bristol, BS16 1QY
United Kingdom
Phone: 44 0117 328 6913
Email: george.jenkinson.2018@bristol.ac.uk

Abstract

Background: In the United Kingdom, women aged 50 to 70 years are invited to undergo mammography. However, 10% of invasive breast cancers occur in women aged ≤45 years, representing an unmet need for young women. Identifying a suitable screening modality for this population is challenging; mammography is insufficiently sensitive, whereas alternative diagnostic methods are invasive or costly. Robotic clinical breast examination (R-CBE)—using soft robotic technology and machine learning for fully automated clinical breast examination—is a theoretically promising screening modality with early prototypes under development. Understanding the perspectives of potential users and partnering with patients in the design process from the outset is essential for ensuring the patient-centered design and implementation of this technology.

Objective: This study investigated the attitudes and perspectives of women regarding the use of soft robotics and intelligent systems in breast cancer screening. It aimed to determine whether such technology is theoretically acceptable to potential users and identify aspects of the technology and implementation system that are priorities for patients, allowing these to be integrated into technology design.

Methods: This study used a mixed methods design. We conducted a 30-minute web-based survey with 155 women in the United Kingdom. The survey comprised an overview of the proposed concept followed by 5 open-ended questions and 17 closed questions. Respondents were recruited through a web-based survey linked to the Cancer Research United Kingdom patient involvement opportunities web page and distributed through research networks’ mailing lists. Qualitative data generated via the open-ended questions were analyzed using thematic analysis. Quantitative data were analyzed using 2-sample Kolmogorov-Smirnov tests, 1-tailed t tests, and Pearson coefficients.

Results: Most respondents (143/155, 92.3%) indicated that they would definitely or probably use R-CBE, with 82.6% (128/155) willing to be examined for up to 15 minutes. The most popular location for R-CBE was at a primary care setting, whereas the most accepted method for receiving the results was an on-screen display (with an option to print information) immediately after the examination. Thematic analysis of free-text responses identified the following 7 themes: women perceive that R-CBE has the potential to address limitations in current screening services; R-CBE may facilitate increased user choice and autonomy; ethical motivations for supporting R-CBE development; accuracy (and users' perceptions of accuracy) is essential; results management with clear communication is a priority for users; device usability is important; and integration with health services is key.
Conclusions: There is a high potential for the acceptance of R-CBE in its target user group and a high concordance between user expectations and technological feasibility. Early patient participation in the design process allowed the authors to identify key development priorities for ensuring that this new technology meets the needs of users. Ongoing patient and public involvement at each development stage is essential.

(J Particip Med 2023;15:e42704) doi:10.2196/42704

KEYWORDS
breast cancer detection; automated diagnosis; breast examination; health care robotics; patient and public involvement; participatory design; user acceptability; mammography; breast cancer

Introduction

Background
Breast cancer is the leading cause of cancer mortality in women worldwide [1]. Almost 11,400 women a year died of breast cancer in the United Kingdom between 2015 and 2017 [2]. However, the mortality rate of breast cancer is falling, with a reduction from 60 per 100,000 in 1989 to 33 per 100,000 in 2017 [2,3]. This trend correlates with the introduction of widespread breast cancer screening using x-ray mammography [4]. In the United Kingdom, mammography is offered to women aged 50 to 70 years through the National Health Service (NHS) Breast Cancer Screening Programme [3]. Screening is estimated to reduce the relative risk of breast cancer mortality by 20% [5] and is linked to many lives saved each year [6].

However, mammography is not suitable for all groups who could benefit from breast cancer screening [7]. For example, 10% of invasive breast cancers occur in younger women (aged <45 years) in the United Kingdom at a rate of 235 per 100,000 [8], a group for whom mammography is not recommended because of its considerably decreased sensitivity in dense breast tissue [7]. This is particularly concerning as young women diagnosed with breast cancer are at higher risk of developing aggressive subtypes and have a poorer prognosis [9]. Mammography is also inappropriate for pregnant women as the low-dose radiation used poses a potential risk during lactation [9]. Furthermore, some women may be unable to tolerate mammography because of pain or discomfort [10]. A breast cancer screening modality that extends services to these groups has the potential to save years of life [11]. Identifying a screening alternative has been challenging. The most effective means of diagnosis (eg, triple assessment and magnetic resonance imaging) are often invasive or costly and unfeasible as screening modalities [12].

A promising alternative is clinical breast examination (CBE) [13,14]. A recent randomized controlled trial of CBE breast cancer screening involving >150,000 women in India demonstrated a 15% reduction in breast cancer–related mortality and a 10% relative risk reduction in the diagnosis of stage-III or stage-IV disease [15]. An overview of systematic reviews assessing the effectiveness of CBE screening identified indirect evidence that CBE has the same effect as mammography when performed well [16]. However, there are several challenges to ensuring that CBE is consistently “well performed.” CBE screening effectiveness may be affected by variations in examination proficiency, training of health care professionals (HCPs), and a lack of standard documentation [17-21].

Recent advances in technology may provide a solution to these challenges. Robotics-assisted procedures have expanded rapidly in recent decades [22,23], and existing literature suggests that health users are increasingly more accepting of artificial intelligence (AI) and machine learning algorithms in cancer screening [24-26]. It is theoretically feasible to create a fully automated robotic CBE (R-CBE) platform by combining soft robotic technology and machine learning algorithms trained by breast specialists. This could offer much-needed standardization of CBE. R-CBE also has the potential to extend screening services to currently underserved groups as it is not reliant on radiation or affected considerably by tissue density. As health policy makers are discussing a risk-stratification approach to breast cancer screening, the cheap and low-risk modality of R-CBE may find further use as part of a strategy to classify the personalized risk level of an individual by measuring physiological properties such as mammographic density [27,28].

The Automated Robotic Examination Intelligent System (ARTEMIS), a novel robotic system for automated CBE, is currently being developed by our research team with support from Cancer Research United Kingdom (CRUK). ARTEMIS aims to combine soft robotic technology with a machine learning platform to allow for fully automated CBE and interpretation of results. The platform could be used by women without direct clinical supervision (Figure 1). A prototype is currently in the early stages of development [29-31]. Although such a platform may be capable of effectively performing and interpreting CBE, the voices of potential users are essential in determining how this should be designed and implemented. Creating technology and a service that is acceptable to end users (and meets their needs) will be crucial in determining the uptake of this type of technology.

Very little published literature is available on the acceptability of intelligent systems that interface directly and independently with users. We identified only 1 study assessing the acceptability of autonomous robotic systems that interface directly with users in health care. This study used robots to perform basic patient assessment tasks (eg, measuring vital signs and inserting intravenous catheters) and concluded that this would be acceptable [32]. We did not identify any publications exploring the acceptability of intelligent robotic services that directly interact with users in cancer screening or diagnostics.
Objectives
This study investigated whether R-CBE is theoretically acceptable to potential users and explored the attitudes, perspectives, and concerns of women regarding the use of intelligent robotic technology in breast cancer screening. It identified key factors that determine whether (and how) the technology would meet the needs of patients, allowing these to be integrated into the prototype design. We adopted the definition of acceptability proposed by Sekhon et al [33]: “a multi-faceted construct that reflects the extent to which people receiving a healthcare intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention.” We conducted a web-based survey of 155 women in the United Kingdom to investigate the following questions: (1) Is there a perceived need for R-CBE? (2) What elements of the R-CBE user interface are most important to women? (3) Is this technology likely to be acceptable to potential users? To the best of our knowledge, this is the first study assessing the acceptability of a fully automated and intelligent patient examination system that interacts directly with users for breast cancer screening.

Methods
A mixed methods approach was used to collect and analyze qualitative and quantitative data from the survey.

Survey Development
The survey consisted of 5 open-ended questions and 17 closed questions with a separate free-text section for respondents to share additional information. A brief overview of the proposed ARTEMIS concept was provided to respondents (Multimedia Appendix 1). This included Figure 1 and a description of how the user might interact with the palpation platform but had no technical details or any information on the accuracy of the device. Our aim was to allow the respondents to freely think about factors that might affect their use of the hypothetical service without imposing any of our priors.

Key constructs were identified based on a review of the health intervention acceptability and health technology literature, our broader knowledge of health technology, and support from the CRUK patient and public involvement specialist team and the London In Vitro Diagnostics Co-operative. We did not identify a fully validated model suitable for our research questions; instead, 2 frameworks were combined with questions selected to cover essential constructs from both. The first was the Theoretical Framework of Acceptability for health care interventions proposed by Sekhon et al [33]. The second was the Unified Theory of Acceptance and Use of Technology developed by Venkatesh et al [34], which has been widely used in research exploring the acceptance of ITs [35]. The resultant key constructs encompassed affective attitudes, perceived effectiveness, ethicality, self-efficacy, effort expectancy, social influence, and facilitating conditions.

The questions were carefully designed to illuminate implicit assumptions and ensure that all key constructs were considered while maintaining an accessible and nonleading language [36]. After a multistage drafting process, the survey was collated and tested on close contacts and members of the associated research department for appropriateness, readability, and ease of use to produce a final draft (Multimedia Appendix 1).

Closed questions allowed us to quantify the overall level of acceptability and desirability of specific features of the service (eg, interface, timing, and preferred location). Thematic analysis of qualitative data provided insights into the quantitative findings. This added richness to our understanding of potential users’ perspectives and attitudes toward the proposed ARTEMIS R-CBE and allowed us to build a more complete picture of acceptability.
Recruitment and Data Collection

Female respondents aged between 20 and 70 years were recruited through a web-based survey linked to the CRUK patient involvement opportunities web page and newsletter and the People in Health West of England and Imperial Human Behaviour and Experience network mailing lists. This nonprobability, voluntary response sampling strategy was chosen because of its quick recruitment rate and ability to serve the exploratory nature of the study. With no hypothesis to test, the aim of the survey was to develop an initial understanding of the needs of the population, and so the bias introduced by self-selection was considered acceptable.

The 15-minute web-based survey was hosted on Qualtrics (Qualtrics International Inc), and 2 attention-check questions were added to ensure that respondents read each question carefully and also to exclude nonhuman (automated) respondents; this resulted in the expulsion of 1 set of responses because the attention questions were answered incorrectly. A further 15 questionnaires were discarded because they were incomplete, including incomplete attention questions, and 3 were discarded because they did not meet the inclusion criteria. This meant that, of 174 responses initiated, 155 (89.1%) completed the survey over 6 weeks between August 2020 and September 2020. Summing the size of each of the mailing lists gives a response rate of 9.26% (155/1674).

Data Analysis

Quantitative analysis was conducted in MATLAB (MathWorks), and differences between groups based on demographics were identified using a 2-sample Kolmogorov-Smirnov test, explored using 1-tailed t tests, and reported where significant (full results available in the data set referenced in the Data Availability section). Pearson correlations were calculated where appropriate to quantify the strength of the associations. CIs were calculated for ranked questions assuming that the preferences were equidistant (1>2>3...).

Qualitative data were analyzed using a method designed around thematic analysis [36]. This allows for detailed exploration of patterns across a data set using a latent approach, with researchers gaining a rich understanding of respondents’ perspectives [36]. Themes were identified after familiarization with the open-text responses. To this end, 2 researchers independently identified a set of key themes within the responses, chosen with relevance to identifying the factors that influenced the respondents’ acceptance of the hypothetical technology. After combining these sets of themes, a single researcher divided each theme into concepts that tightly grouped responses within each theme. Salient ideas from these grouped concepts were then extracted to describe the outcomes of the responses as a whole.

The raw data are available from the source provided in the Data Availability section at the end of this paper.

Ethics Approval

The study received ethics approval from the Imperial College Research Ethics Committee (20IC6129).

Results

Quantitative Results

Demographics

The average age of the respondents was 49.8 (SD 12.7; range 21-70) years. “White” ethnic (142/155, 91.6%) and university-educated (119/155, 76.8%) backgrounds were overrepresented among survey respondents. Our study population also had a higher incidence of personal history of breast cancer (28/155, 18.1%) compared with the general adult population (4.46% [8]). Respondents were overwhelmingly in favor of screening programs (143/155, 92.3%) and the use of technology in health care (146/155, 94.2%). The demographic data are summarized in Table 1, and attitudes toward screening and technology in health care in general are summarized in Table 2.
### Table 1. Respondent demographics (N=155).

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>21-44</td>
<td>54 (34.8)</td>
</tr>
<tr>
<td>45-59</td>
<td>48 (31)</td>
</tr>
<tr>
<td>60-70</td>
<td>51 (32.9)</td>
</tr>
<tr>
<td>Did not complete</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White British</td>
<td>121 (78.1)</td>
</tr>
<tr>
<td>Other White</td>
<td>21 (13.5)</td>
</tr>
<tr>
<td>Black African</td>
<td>4 (2.6)</td>
</tr>
<tr>
<td>Indian</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>White and Black African</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Pakistani</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>White and Black Caribbean</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Chinese</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td><strong>Highest qualification</strong></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>119 (76.8)</td>
</tr>
<tr>
<td>Vocational qualification (ONC, BTEC, or NVQ)</td>
<td>13 (8.4)</td>
</tr>
<tr>
<td>A-Less (or equivalent)</td>
<td>13 (8.4)</td>
</tr>
<tr>
<td>GCSE or O-Levels (or equivalent)</td>
<td>10 (6.5)</td>
</tr>
<tr>
<td><strong>History of diagnosis of cancer</strong></td>
<td></td>
</tr>
<tr>
<td>Any (including breast cancer)</td>
<td>42 (27.1)</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>28 (18.1)</td>
</tr>
</tbody>
</table>

aONC: Ordinary National Certificate.
bBTEC: Business and Technology Education Council qualification.
cNVQ: National Vocational Qualifications.
dGCSE: General Certificate of Secondary Education.

### Table 2. Respondents’ attitudes toward breast cancer screening and technology in health care (N=155).

<table>
<thead>
<tr>
<th>Questions and responses</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you think routine cancer screening tests are a good idea?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>143 (92.3)</td>
</tr>
<tr>
<td>No</td>
<td>4 (2.6)</td>
</tr>
<tr>
<td>Don’t know</td>
<td>8 (5.2)</td>
</tr>
<tr>
<td>What do you think of increased use of new technology in health care?</td>
<td></td>
</tr>
<tr>
<td>Very bad idea</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Bad idea</td>
<td>5 (3.2)</td>
</tr>
<tr>
<td>Good idea</td>
<td>48 (31)</td>
</tr>
<tr>
<td>Very good idea</td>
<td>98 (63.2)</td>
</tr>
<tr>
<td>Don’t know</td>
<td>5 (3.2)</td>
</tr>
</tbody>
</table>
Overall Opinion Toward the Device

Provided the R-CBE was as good as an HCP, 92.3% (143/155) of respondents said that they would either “definitely” (104/155, 67.1%) or “probably” (39/155, 25.2%) use an R-CBE service if it were offered. In comparison, 89.7% (139/155) of respondents said that they would “definitely” (92/155, 59.4%) or “probably” (47/155, 30.3%) use a service offering CBE by a trained HCP (Figure 2). This indicates that the answers to the 2 questions were similar, with a slight preference for R-CBE (2-sample Kolmogorov-Smirnov test; \( P = .40 \)). Willingness to use an R-CBE service was moderately correlated with respondents’ likelihood of using new technology in general \( (r_{155}=0.4014; P<.001) \).

Respondents were asked to indicate which factors would make them more likely to use R-CBE. The most popular option was receiving a “faster referral to specialist breast services” if required (144/155, 92.9% of respondents selected this option) and being able to drop in and use the device without an appointment (108/155, 69.7% of respondents). Other factors that influenced anticipated use are shown in Table 3.

Figure 2. Overall opinion of the device. This demonstrates that the idea of a robotic system appeals to some respondents more so than the status quo. CBE: clinical breast examination.

Table 3. Factors to improve uptake, which provides insights into the respondents’ understanding of how a robotic system might best be of benefit to them \((N=155)\).

<table>
<thead>
<tr>
<th>What would make you more likely to use R-CBE(^a)</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster referral to a specialist</td>
<td>93 (60)</td>
</tr>
<tr>
<td>Drop-in appointments</td>
<td>70 (45.2)</td>
</tr>
<tr>
<td>Knowing what to expect before the appointment</td>
<td>68 (43.9)</td>
</tr>
<tr>
<td>My GP(^b) seeing the results</td>
<td>61 (39.4)</td>
</tr>
<tr>
<td>Confidential results</td>
<td>59 (38.1)</td>
</tr>
<tr>
<td>More technical information</td>
<td>39 (25.2)</td>
</tr>
<tr>
<td>Information on data protection</td>
<td>36 (23.2)</td>
</tr>
</tbody>
</table>

\(a\)R-CBE: robotic clinical breast examination.\
\(b\)GP: general practitioner.

Device Features

The respondents favored the use of soft (rather than hard) robotic parts for the aspects of the device that would be in contact with their skin. Device features considered to be of most importance were availability of information on access to support from an HCP, appointment availability, cleanliness, and regular updates on examination progress throughout the procedure. The results are presented in Table 4.

A comparative analysis of the age groups revealed 3 significant differences. Each respondent scored a selection of features on a scale of 1 to 5. The age group of >60 years (the oldest) considered ease of appointment availability to be less important compared with the 2 younger age groups (>60 years vs 45 to 59 years: \(mean difference [MD]=0.37\) and \(P=.02\); >60 years vs <45 years: \(MD=0.43\) and \(P=.04\)). Conversely, the age group of <45 years considered it less important to be able to adjust the speed of the device (<45 years vs 45 to 59 years: \(MD=0.59\) and \(P=.04\); <45 years vs >60 years: \(MD=0.87\) and \(P=.001\)) or for the device to have disposable parts (<45 years vs 45 to 59 years: \(MD=0.89\) and \(P=.002\); <45 years vs >60 years: \(MD=0.70\) and \(P=.02\)).
### Table 4. Relative importance of device features. “On a scale from 1 (not at all important) to 5 (essential), how important is it that...”

<table>
<thead>
<tr>
<th>Feature</th>
<th>Score, mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The device provides links to support from HCPs(^a)</td>
<td>4.26 (4.12-4.39)</td>
</tr>
<tr>
<td>Appointments are easily available</td>
<td>4.16 (4.04-4.28)</td>
</tr>
<tr>
<td>Information about the cleaning of the booth is available</td>
<td>4.10 (3.96-4.25)</td>
</tr>
<tr>
<td>The examination provides constant updates</td>
<td>4.08 (3.96-4.21)</td>
</tr>
<tr>
<td>The device is close to home or work</td>
<td>3.63 (3.49-3.78)</td>
</tr>
<tr>
<td>Parts of the device that are in contact with the skin are disposable</td>
<td>3.34 (3.14-3.55)</td>
</tr>
<tr>
<td>I am able to adjust the speed of the device’s parts that are in contact with the skin</td>
<td>2.76 (2.58-2.94)</td>
</tr>
</tbody>
</table>

\(^a\)HCP: health care professional.

### Location

Most respondents (130/155, 83.9%) preferred the booth to be located at a site associated with health care. The most popular location was at a general practitioner surgery, which generally provides point-of-contact care and triage between patients and specialist health services in the United Kingdom, followed by “inside a pharmacy.” Options not associated with health care (such as at a shopping center or in the workplace) were less popular. This difference was statistically significant. Location preference is shown in Table 5. The age group of >60 years favored the shopping center more compared with the other age groups (>60 years vs 45 to 59 years: MD=0.50 and \(P=0.02\); >60 years vs <45 years: MD=0.57 and \(P=0.008\)).

### Table 5. Location preference.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Option</th>
<th>Rank, mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GP(^a) surgery</td>
<td>1.41 (1.23-1.50)</td>
</tr>
<tr>
<td>2</td>
<td>Pharmacy</td>
<td>2.15 (1.97-2.20)</td>
</tr>
<tr>
<td>3</td>
<td>Shopping center</td>
<td>3.77 (3.47-3.82)</td>
</tr>
<tr>
<td>4</td>
<td>Work</td>
<td>3.84 (3.52-3.91)</td>
</tr>
</tbody>
</table>

\(^a\)GP: general practitioner.

### Length of Examination

Most respondents (153/155, 98.7%) were willing to be examined for up to 10 minutes, 82.6% (128/155) were willing to be examined for up to 15 minutes, and 56.8% (88/155) were willing to be examined for 20 minutes. Interestingly, only 22.6% (35/155) of the respondents considered the time taken to carry out the examination to be “Quite important” (34/155, 21.9%) or “Essential” (1/155, 0.6%). When asked to rate “how important is it that the examination does not take longer [than the duration indicated]” on a scale of 0 (not at all important) to 5 (essential), the mean rating was 1.46. However, respondents who preferred a shorter examination duration were statistically more likely to report that it was important that the examination last no longer than they had indicated (\(r_{153}=0.53; P<0.001\)). These results are shown in Figures 3 and 4.
Communication of Results

Most users (128/155, 82.6%) preferred to receive information directly from the device, either displayed on the screen with a printout (mean rank 2.26) or received via email (mean rank 2.47). These options were statistically significantly more popular than the results being emailed to their physician first. This was true both in the case of a normal (mean rank 5.02) and an abnormal (mean rank 3.72) result. In the event of an abnormal result, the option “email to my doctor first” increased in preference (from sixth to fourth in the average rank) but remained comparatively unpopular. The most popular option for results communication was through a combination of written information and pictures (mean rank 1.62). Respondents without a university diploma or equivalent ranked seeing their results on-screen without a printout significantly higher than those with a university diploma or equivalent (healthy: $MD=0.57$ and $P=0.08$; abnormal: $MD=0.89$ and $P=0.02$). This suggests that the level of education may be an important discriminant when considering how results are communicated. Respondents highly valued the inclusion of information on appropriate follow-up and alternative explanations for identified abnormalities. The results are summarized in Tables 6-8.

It is worth noting that 6.5% (10/155) of respondents used the open-text “other” option to indicate that they would want to receive results from an HCP and not from the R-CBE device itself. All respondents (155/155, 100%) ranked this as their number 1 preference.
Table 6. Information receipt preferences.

<table>
<thead>
<tr>
<th>Option</th>
<th>No referral advised, mean rank (95% CI)</th>
<th>Referral advised, mean rank (95% CI)</th>
<th>Rank change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediately on-screen+printout</td>
<td>2.26 (2.02-2.50)</td>
<td>2.41 (2.20-2.62)</td>
<td>−0.15</td>
</tr>
<tr>
<td>Emailed later</td>
<td>2.47 (2.26-2.68)</td>
<td>2.66 (2.49-2.83)</td>
<td>−0.19</td>
</tr>
<tr>
<td>Immediately on-screen</td>
<td>2.76 (2.57-2.96)</td>
<td>3.30 (3.06-3.53)</td>
<td>−0.52</td>
</tr>
<tr>
<td>SMS text message</td>
<td>4.10 (3.90-4.29)</td>
<td>4.61 (4.37-4.75)</td>
<td>−0.46</td>
</tr>
<tr>
<td>Posted later</td>
<td>4.24 (4.07-4.41)</td>
<td>4.25 (4.05-4.44)</td>
<td>−0.01</td>
</tr>
<tr>
<td>Emailed to physician first</td>
<td>5.02 (4.82-5.21)</td>
<td>3.72 (3.43-4.00)</td>
<td>+1.31</td>
</tr>
<tr>
<td>Other</td>
<td>7.40 (7.11-7.68)</td>
<td>6.40 (6.16-6.63)</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 7. Information display preferences.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Option</th>
<th>Rank, mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Written and pictures</td>
<td>1.62 (1.48-1.76)</td>
</tr>
<tr>
<td>2</td>
<td>Interactive app</td>
<td>2.02 (1.84-2.20)</td>
</tr>
<tr>
<td>3</td>
<td>Written only</td>
<td>2.53 (2.38-2.68)</td>
</tr>
<tr>
<td>4</td>
<td>Verbal summary</td>
<td>3.785 (3.6-3.97)</td>
</tr>
</tbody>
</table>

Table 8. Information content preferences (N=155).

<table>
<thead>
<tr>
<th>What information would you like included in your results</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details on follow-up when referral is recommended</td>
<td>153 (98.7)</td>
</tr>
<tr>
<td>How to book a future R-CBE(^a) appointment</td>
<td>151 (97.4)</td>
</tr>
<tr>
<td>Other causes for an “abnormal” finding</td>
<td>122 (78.7)</td>
</tr>
<tr>
<td>Links to emotional support</td>
<td>108 (69.7)</td>
</tr>
</tbody>
</table>

\(^a\) R-CBE: robotic clinical breast examination.

Qualitative Results

Overview

Qualitative analysis of the free-text responses identified the following seven superordinate themes with respect to R-CBE: (1) women perceived that R-CBE has the potential to address limitations in current screening services, (2) R-CBE may facilitate increased user choice and autonomy, (3) ethical motivations for supporting R-CBE development, (4) accuracy (and users’ perceptions of accuracy) is a priority, (5) results management with clear communication is a priority for users, (6) integration with health services is key, and (7) device usability is important. These themes are summarized in the following sections. Quotes from the respondents illustrating the themes are shown in Table 9.
Table 9. Themes from thematic analysis with supporting quotes.

<table>
<thead>
<tr>
<th>Theme and concept</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R-CBE has the potential to address limitations in current services</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Provides reassurance | “I worry about my breast health. It would be reassuring to be able to check for irregularities.” [Respondent 90]  
“...to be able to regularly monitor for something like breast cancer would give me peace of mind.” [Respondent 008] |
| Reluctance to “waste” physicians’ time | “I find [breast examination] difficult to do myself and don’t like to take up doctors time very often.” [Respondent 032]  
“I would also like regular check-ups and understand GPs need to prioritise other appointments.” [Respondent 097] |
| Negative experiences with mammography | “When I have a mammogram it really hurts me. I often say that the machine is like torture.” [Respondent 129]  
“I would welcome any solution that is pain free.” [Respondent 129] |
| Embarrassment during clinical examination | “Every time I see breast screening on TV there is a picture of women with a completely naked top half. This makes me feel very uncomfortable and puts me off screening.” [Respondent 053]  
“I would rather have a machine examine my breasts than a doctor. It would eliminate the feeling of embarrassment.” [Respondent 036]  
“...lack of human contact may encourage more women to use it.” [Respondent 087] |
| Anxiety associated with awaiting results | “Every time I have a mammogram, I panic for 2-3 weeks waiting for the results.” [Respondent 134]  
“If results are available immediately then that’s better than waiting for test results and stops stress and anxiety.” [Respondent 106]  
“The possibility of having instant results is amazing.” [Respondent 142] |
| **R-CBE may facilitate increased user choice and autonomy** | |
| Choice over appointment time, frequency, and location | “...it’s is [sic] more convenient if you have a bigger choice over appointment times.” [Respondent 039]  
“Freedom to choose when to use the device.” [Respondent 029]  
“...hopefully accessibility (location/appointments) are easier than going to the GP” [Respondent 049]  
“...this would allow more frequent checks.” [Respondent 147] |
| Increased sense of autonomy | “...the opportunity to be firmly in control of ones [sic] own health concerns is appealing.” [Respondent 143]  
“I believe the autonomy of this device may encourage more people to come forth for screening.” [Respondent 132] |
| **Ethical motivations for supporting R-CBE** | |
| Support for population screening in general | “I like to spread the word about health screening, it’s very important to look after your health.” [Respondent 053]  
“I would support anything that encourages people to be tested.” [Respondent 054] |
| Potential to increase access for underserved populations | “I think it is very important that women can have regular breast examinations that start at a younger age that [sic] mammograms!” [Respondent 007]  
“Digital Automation seems to be one way of improving life chances for black Cancer patients like myself.” [Respondent 144]  
“[a family member] has a learning difficulty, is deaf and is a wheelchair user. It has not been possible for her to have the benefit of regular breast screening. I am hopeful that this new device will help women like her in the future.” [Respondent 143] |
| Reduced burden on the NHS | “Technology is advancing and the population is growing. Using this technology in health care will help to free up our medical staff so that they can use their much needed skills working it [sic] areas where only human intervention is possible.” [Respondent 063]  
“It would seem to be an efficient screening tool that allows precious medically trained staff to do other jobs a machine cannot do.” [Respondent 111] |
| **Accuracy, and users’ perception of accuracy, is a priority** | |
| R-CBE is only acceptable if users are convinced of its accuracy | “I think it is a very good idea...provided there is a definite level of assured accuracy.” [Respondent 062]  
“I would use it if I had confirmation that results are accurate.” [Respondent 054] |
Theme and concept | Quotes
--- | ---
Factors that increase confidence in accuracy | • "...if the technology and device is proven through appropriate clinical studies." [Respondent 155]
• "...how often it gets the diagnosis right, how often it gets it wrong." [Respondent 022]
• "MHRA [Medicines and Health care products Regulatory Agency] approval." [Respondent 041]
• "...some sort of checking procedure e.g. [sic] every 50th person is called in for manual checks." [Respondent 054]

Suitable results management with sensible communication

Sensitivity concern about receiving results from R-CBE | • "I worry about the emotional impact of an abnormal result being given via automated means.” [Respondent 006]
• "I think a human can usually be gentler with the feelings of patients.” [Respondent 022]

Rapid results are preferred | • "[I] wouldn’t want doctor involvement to delay my getting the result.” [Respondent 042]
• "I would much prefer the results at the time of the test.” [Respondent 134]

Factors that optimize user experience when receiving results | • "I’d want to know more about what an ‘abnormal’ result might mean—does it definitely mean cancer, or could it mean something else?” [Respondent 100]
• "If there is an abnormal result it will cause an amount of worry and anxiety and so any additional information that can be provided alongside the results such as emotional support and links to further information would be really useful.” [Respondent 142]
• "[if] the results are abnormal...an automatic urgent appointment should be made by the GP straight away.” [Respondent 079]

Confidentiality and privacy are essential | • "...all the physical privacy and data privacy issues [need to be] well thought through.” [Respondent 059]
• "To be screened in a booth, it would have to be entirely 100% privacy proof, confidential, and safe.” [Respondent 134]

Integration with health services is key

High trust placed in the NHS | • "I would use a machine if it ran in tandem with NHS services.” [Respondent 022]
• "If it’s recommended by my GP or other relevant HCP.” [Respondent 141]
• "I would expect it to complement other services not replace them.” [Respondent 145]

Geographic proximity to other health services | • "I think I would feel more comfortable if the service was in a health care setting (e.g. GP/pharmacy), rather than in a more public space (e.g. work).” [Respondent 014]
• "I think the location should be somewhere linked to medical care/support—even if just near a first aider’s office.” [Respondent 146]

Device usability is important

Clear instructions required | • "...if the instructions are fool-proof I think I could manage it.” [Respondent 063]
• "[needs] clear and understandable for everyone.” [Respondent 082]
• "A video demo would be helpful to maybe watch before attending.” [Respondent 148]

Clear plan for managing technical difficulties | • "I might need a little reassurance that the machine wasn’t going to run amok.” [Respondent 149]
• "My only reservation was if it went wrong and either used the wrong pressure or wouldn’t unclamp from the breast.” [Respondent 055]
• "Where to get help if the device didn’t work or stopped working during examination.” [Respondent 145]
• "...a panic or immediate stop function [with] the ability to cancel and walk away.” [Respondent 151]

Women Perceive That R-CBE Has the Potential to Address Limitations in Current Screening Services

The limitations of current breast cancer screening services were raised frequently, and respondents perceived that R-CBE has the potential to address some of these limitations. “Check-ups” could provide regular reassurance lacking in current services. Respondents recognized that they could regularly self-examine (but lacked confidence to do so) or request regular examinations from a health practitioner (but did not want to waste the physician’s time). Pain associated with mammography was the most frequently cited limitation of breast cancer screening. Many respondents assumed that soft robotics would be more comfortable than a mammogram. R-CBE could also reduce the embarrassment of being seen unclothed by an HCP during mammography or CBE. Some respondents believed that a fully automated service that reduced this embarrassment was preferable to direct human involvement. Long waiting times to...
receive screening results were associated with anxiety, and the possibility of receiving rapid results from automated technology was highly appealing. This theme reflects the potential of R-CBE to address limitations in current services.

**R-CBE May Facilitate Increased User Choice and Autonomy**

R-CBE may be “more convenient” than other screening services, offering a wider choice of appointment times, location, and the frequency with which the service could be accessed. This increased choice over where and when, combined with the opportunity to complete screening without input from an HCP, was appealing and provided a sense of autonomy and control.

**Ethical Motivations for Supporting R-CBE Development**

Some respondents viewed R-CBE favorably on an ethical basis. For example, respondents suggested R-CBE (with the potential to be a convenient and accessible service) could increase screening among traditionally underserved populations such as young women, ethnic minorities, or people with disabilities. There was a desire to extend screening and cancer prevention on a population basis, irrespective of the modality, and strong support for the NHS. Respondents indicated that they would accept R-CBE if it reduced the burden on the NHS and HCPs. This reflects an underlying assumption that an automated device screening service would reduce the burden on the NHS. This assumption is explored further in the Discussion section. This theme indicates support for the R-CBE concept based on the respondents’ broader attitudes and ethical beliefs.

**Accuracy, as well as Users’ Perception of Accuracy, Is a Priority**

Acceptance of R-CBE was conditional, and respondents identified several factors required for R-CBE to be trustworthy. Chief among these was accuracy. Unsurprisingly, the requirement that the device have high levels of accuracy was mentioned by most respondents (132/155, 85.2%) unprompted. There was no clear required accuracy threshold. Some respondents wanted to see a service that was “as good as a mammogram,” others wanted to see a service “as good as a GP,” and others still “would use the device on the condition that it was better than a doctor.” However, there was a consensus that users should be provided with enough information to make their own informed decision as to whether R-CBE is accurate enough. Respondents suggested that users be given information on the sensitivity and specificity, ongoing monitoring of device performance, clinical trials completed, and regulatory approval to optimize trust. To be trustworthy, R-CBE must be highly accurate, and salient understandable information on how this accuracy is determined must be made available.

**The Need for Suitable Results Management With Sensitive Communication**

Communication of results in a sensitive manner was a key priority. Receiving screening results is anxiety-inducing, and the responses indicated that this is particularly true for technology-based services. Some respondents expressed concern about the ability of R-CBE to do this in a sufficiently sensitive manner. A small number of respondents felt that direct human involvement was essential in the event of an abnormal result. They felt strongly about this and described the idea of receiving an abnormal result from an automated device as “cold,” “impersonal,” and “abhorrent.” However, more respondents reported that rapid availability of results outweighed this disadvantage. Options for optimizing direct R-CBE results delivery were identified. These included ensuring an efficient follow-up process, providing information on possible causes of an abnormal result (options other than malignancy), and providing guidance on where users could access support if needed. It was also important to respondents that results management be private and confidential and that detailed information on data storage be available.

This theme illustrates the need for efficient, sensitive, private, and secure processes for managing results that place users first. Providing sufficient information to service users may optimize the experience and minimize the anxiety associated with receiving results.

**Integration With Health Services Is Key**

Along with timely follow-up of abnormal results, functional integration with the health service was highly valued. Adequate integration with the health service appeared to increase user confidence in the new technology. A high degree of trust was placed in the NHS, and integration with this would lend credibility to R-CBE. It was important that the new technology be an adjunct to existing services without reducing access to general practitioners or current NHS services. Geographic proximity to existing health services was also viewed positively as respondents perceived that this could improve integration and access to support. The trustworthiness of R-CBE appears to depend not only on the device itself but also on the extent to which it is integrated into the existing health system.

**Device Usability Is Important**

Acceptability was conditional on R-CBE being easy to use. People must also be confident that they can use the device without compromising accuracy. The importance of clear instructions was highlighted; providing a short instructional video was a suggested method of ensuring this. There was also a degree of anxiety regarding the possibility of malfunctions. Respondents wanted a clear procedure for dealing with technical difficulties. Suggestions for this included an emergency stop function and a process for calling for assistance. Clear instructions, a plan for malfunctions, and an emergency stop button would provide peace of mind and respect women’s autonomy by giving them control over the examination.

**Discussion**

**Principal Findings**

Responses were generally positive for a potential R-CBE service that is at least equivalent to a nonrobotic alternative. The overwhelming majority of respondents reported that they would use R-CBE screening if it were offered. Respondents recognized the potential of R-CBE to address an unmet need in current screening services by providing regular reassurance, reducing interpersonal embarrassment, reducing screening-associated pain, improving appointment availability, and offering rapid
results. All these are barriers to screening uptake recognized in the existing literature [37].

The survey showed that a high level of sensitivity and specificity of this technology is an essential factor for acceptability. User acceptance in our survey was dependent on R-CBE being a highly accurate system.

The results of this study also complement the existing literature on AI diagnostics, which suggests that the public has a high level of trust in computerized decision-making in health care and that AI in cancer screening is increasingly accepted [24-26].

The acceptance of R-CBE was qualified. Our results complement the existing literature [38] by identifying high levels of trust as an essential property for the uptake of robotic and automated systems. Our data identified factors that are necessary for an R-CBE service to be considered trustworthy. Key among these are accuracy, usability, and communication. Respondents’ concerns regarding the lack of human connection, data privacy, and regulation of new health technology echoed similar concerns identified in a recent study exploring the public perception of AI mammography reading [26].

Our results indicate that most users are likely to accept autonomous screening if there is a well-established, efficient process for follow-up with a clinician if needed. This agrees with studies to date indicating that people are more accepting of intelligent systems working symbiotically with physicians or HCPs [24,39] but remain ambivalent about those that function independently [25].

This study provides important information to guide decision-making on R-CBE development, determine its viability as an investment, and inform our understanding of public attitudes toward intelligent health technology in cancer screening. Crucially, our results indicated a significant concordance between what is technically feasible and what is acceptable to users. For example, most respondents (128/155, 82.6%) were willing to be examined for up to 15 minutes and were also willing to receive results directly from ARTEMIS (in some format) rather than from an HCP. Research suggests that it is feasible to create an automated R-CBE service based on these acceptability characteristics [29-31].

Limitations and Future Directions

From these results, we believe that R-CBE may offer a more patient-focused option that has the potential to increase screening uptake provided it can perform examinations with sufficient sensitivity and specificity.

To develop technologies seeking to provide the service of R-CBE or similar, these results provide appropriate targets to be met when evaluating their expected acceptability. For example, several respondents supported R-CBE because it would reduce the burden on the NHS and free up time for HCPs. Although early detection and intervention could reduce progression to advanced disease (and, therefore, reduce the treatment burden on the NHS), this assumption is only valid if R-CBE detects early disease and allows for early intervention without overdiagnosis or excessive referrals to primary or specialist services.

As an investigative survey, the sample size was comparatively small, and the skewed distribution of demographic groups within the sample means that it was insufficiently powered to detect nuanced differences between them. A larger sample size with a demographic distribution representative of the wider population would be needed to identify whether the subtle differences in preference between demographic groups in this study are statistically significant and externally valid in the general population.

The demographics of the respondents were also not representative of the UK population. First, Black and minority ethnic groups were underrepresented. The data may not accurately capture the needs, thoughts, attitudes, and perspectives of these demographics. This is of particular concern as these groups are at an elevated risk of breast cancer and face the greatest barriers to screening [40,41]. Reaching these groups in future research is essential as they may benefit substantially from widening screening. Achieving this is likely to require targeted methods.

In addition, over three-quarters of our sample (119/155, 76.8%) had a degree-level education. Jonmarker et al [25] found a significant association between level of education and level of trust in technology. This is reflected in the very high levels of trust in technology reported in our sample. This reduces the generalizability of our results, with survey respondents being more likely to find R-CBE acceptable than the general population. The non–probability sampling used in this study may also introduce selection bias—it is possible that women who had a history of engaging with existing breast cancer screening programs were more likely to answer the survey, which might have contributed to overestimation of the acceptability of R-CBE screening. The particular method of electronic survey requires respondents to have ready access to a compatible device connected to the internet and be literate at using it, inherently excluding those who do not fulfill both criteria.

The ARTEMIS R-CBE is currently in development (part of this system is described in the study by Jenkinson et al [31]); the responses relate to a theoretical service. Further research will be needed to establish the acceptability of the specific service among users as development continues, as well as an assessment of its cost and accuracy. Future research would benefit from a larger and more diverse sample size that better represents the population. Our team is currently undertaking further qualitative research via focus groups to better understand the requirements of trustworthy and acceptable R-CBE and automated breast cancer screening more generally. Despite the limitations outlined previously, the survey data allowed us to identify key priorities among potential users and provide valuable information for the research team. These findings may provide insights for others working in automated health technology development, particularly for cancer screening.

Conclusions

R-CBE holds promise as a new modality of breast cancer screening. It could address limitations in current screening services, increase screening uptake, and provide a more patient-focused service. This investigative survey demonstrated
that there is potential for high levels of acceptability of R-CBE among its target user group and a high concordance between user expectations and technological feasibility. However, the acceptability of R-CBE is conditional on users being confident that it is accurate, easy to use, able to communicate results sensitively, and well integrated with health services. These findings will contribute directly to prototype development and will be of interest to other researchers developing automated cancer screening and related health technologies. This study highlights the fact that the development of new technologies raises ethical and practical issues. The importance of public and patient involvement in health technology development to address these issues should not be underestimated. Patient and public involvement at each stage of development will be key to ensure that any future service meets the needs of the public.

Acknowledgments
The authors would like to acknowledge the support of the Cancer Research UK Patient and Public Involvement team for their invaluable advice. They would also like to acknowledge People in Health West of England, who promoted our survey. GPJ and AT would like to thank Miss Sasirekha Govindarajulu for providing invaluable experience and insight into breast cancer clinics. Finally, the authors would like to acknowledge the respondents for their time and participation. This work was supported by project Automated Robotic Examination Intelligent System, Cancer Research UK (grant C24524/A30038), and the Engineering and Physical Sciences Research Council (grant EP/R513179/1).

Data Availability
The data sets generated for this study can be found on GitHub [42].

Conflicts of Interest
None declared.

Multimedia Appendix 1
Questionnaire. [DOCX File, 203 KB - jopm_v15i1e42704_app1.docx ]

References


42. geojenks - Acceptability Survey. GitHub. URL: https://github.com/geojenks/Acceptability_Survey [accessed 2023-03-13]

Abbreviations

AI: artificial intelligence
ARTEMIS: Automated Robotic Examination Intelligent System
CBE: clinical breast examination
CRUK: Cancer Research United Kingdom
HCP: health care professional
MD: mean difference
NHS: National Health Service
R-CBE: robotic clinical breast examination

©George P Jenkinson, Natasha Houghton, Nejra van Zalk, Jo Waller, Fernando Bello, Antonia Tzemanaki. Originally published in Journal of Participatory Medicine (https://jopm.jmir.org), 03.04.2023. 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 Journal of Participatory Medicine, is properly cited. The complete bibliographic information, a link to the original publication on https://jopm.jmir.org, as well as this copyright and license information must be included.
Development of a Family-Centered Communication Tool for Kidney Health in Premature Infants: Qualitative Focus Group Study Using Human-Centered Design Methodology

Michelle C Starr1, MPH, MD; Samantha Wallace2, MSc; Courtney Moore1, BFA; Brandon Cockrum3, MFA; Bridget Hawryluk3, MFA; Aaron Carroll2, MSc, MD; William Bennett Jr3, MSc, MD

1Division of Pediatric Nephrology, Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, United States
2Center for Pediatric and Adolescent Comparative Effectiveness Research, Indiana University School of Medicine, Indianapolis, IN, United States
3Research Jam, Indiana Clinical and Translational Sciences Institute, Indianapolis, IN, United States

Corresponding Author:
Michelle C Starr, MPH, MD
Division of Pediatric Nephrology
Department of Pediatrics
Indiana University School of Medicine
410 W 10th Street
Suite 2000A
Indianapolis, IN, 46038
United States
Phone: 1 317 278 9817
Email: mcstarr@iu.edu

Abstract

Background: Premature infants are at increased risk of kidney-related complications, including acute kidney injury (AKI) and chronic kidney disease (CKD). The risk of CKD in prematurely born infants is underrecognized by health care teams and caregivers. Understanding how to communicate the risk of CKD to caregivers is essential for longitudinal clinical follow-up and adherence.

Objective: This study aimed to determine family caregiver attitudes toward kidney health and risk communication during a neonatal intensive care admission. We also sought to understand caregiver preferences for the communication of information surrounding the risk of CKD in premature infants.

Methods: We augmented standard qualitative group sessions with human-centered design methods to assess parent preferences and clinician perspectives. Caregivers recruited had a prematurely born child who spent time in the neonatal intensive care unit at Riley Hospital for Children in Indianapolis, Indiana, and experienced AKI or another kidney complication, which put them at risk for future CKD. We used a variety of specific design methods in these sessions, including card sorting, projective methods, experience mapping, and constructive methods.

Results: A total of 7 clinicians and 8 caregivers participated in 3 group sessions. Caregivers and clinicians readily acknowledged barriers to and drivers of long-term kidney monitoring as well as opportunities for communication of the risk of long-term kidney disease. Caregivers’ primary concerns were for both the type and depth of information conveyed as well as the time at which it was communicated. Participants emphasized the importance of collaboration between the hospital care team and the primary care provider. Participant input was synthesized into several prototype concepts and, ultimately, into a rough prototype of a website and an informational flyer.

Conclusions: Caregivers of premature infants are open to communication about kidney health during their neonatal admission. The next phase of this work will translate caregivers’ preferences into family-centered communication tools and test their efficacy in the neonatal intensive care unit.

(J Particip Med 2023;15:e45316) doi:10.2196/45316

KEYWORDS

qualitative research; patient-reported outcomes; neonates; chronic kidney disease; human-centered design; acute kidney injury; kidney health
**Introduction**

Premature infants are at high risk of kidney-related complications, including acute kidney injury (AKI) and chronic kidney disease (CKD) [1]. AKI is common in premature infants, occurring in between 20% and 40% of infants, depending on the patient population studied [2,3]. Premature infants with AKI have higher rates of mortality and longer hospital stays [2,3]. The risk of kidney-related complications in premature infants does not disappear after the neonatal admission. Studies in prematurely born children show a 4-fold increase in CKD during childhood and adolescence [4-9]. While likely multifactorial, one explanation for this increased CKD risk is that premature infants are born with a decreased number of nephrons due to their early delivery [10]. Furthermore, the extracellular environment (including the use of nephrotoxic medications and perinatal stressors) may not be amenable to proper nephron development [11]. Even children with normal kidney function but a history of AKI have a 10 times higher risk of developing kidney failure before the age of 40 years [12,13]. Thus, as more critically ill infants survive and live into adulthood, the impact of kidney health on premature infants is a significant long-term concern.

Communication surrounding kidney health to families, specifically focusing on the risk of CKD, is essential in empowering families and ensuring longitudinal clinical follow-up and monitoring. Studies show that kidney health, including the diagnosis of AKI and the risk of CKD in prematurely born infants, is underrecognized by health care teams [14,15]. While there are no established best practices for communication in the neonatal intensive care unit (NICU), families of premature infants report a desire for direct and concise communication during their NICU stay, focusing on the most urgent or immediate clinical concerns [16,17].

There have been no studies which evaluate kidney-specific health communication with families. The purpose of this study was to fill this gap by evaluating caregiver attitudes toward kidney health and CKD risk communication as well as caregiver preferences for communication of information surrounding the risk of CKD. This study serves as the first step in the development of a family-centered tool to improve communication about kidney health in premature infants.

**Methods**

**Overall Approach**

In collaboration with Research Jam and the Indiana Clinical and Translational Science Institute’s Patient Engagement Core, we conducted 2 phases of group sessions using qualitative focus group methodology augmented by human-centered design methods (Multimedia Appendix 1). Human-centered design, which is increasingly used within health care, is an iterative design process where stakeholders most closely affected by the problem or solution are engaged in developing the solution [18,19].

Sessions were facilitated by 4 research specialists using human-centered design research methods. Sessions were held virtually through Zoom (Zoom Video Communications), lasted approximately 120 minutes each, and were recorded and transcribed for analysis. All sessions used activities to engage clinicians and caregivers to better understand caregiver perspectives on communication surrounding kidney disease as a first step in the co-design of a kidney disease communication tool [20,21]. Activities were open-ended, allowing for a wide range of responses to minimize bias and for families to be as open and truthful as possible about their preferences. Sessions began with warm-up activities to encourage participation and collaboration [22]. We then used specific generative activities (eg, empathy mapping, detailed below) designed to encourage study participants to express their thoughts and feelings and constructive methods to help with concept development [21]. All sessions used Miro Whiteboard (Miro) [23], a collaborative whiteboard platform which the group facilitator used to document and visualize responses in real-time for the group.

**Recruitment, Subjects, and Study Setting**

Stakeholders included clinicians and caregivers. Clinicians were from across the United States and cared for prematurely born children who spent time in the NICU. This included physicians, nurses, and nurse practitioners trained in pediatric nephrology, general pediatrics, and perinatal and neonatal medicine, all of whom were approached and recruited by the principal investigator.

Caregivers were recruited who had prematurely born children who spent time in the NICU at Riley Hospital for Children in Indianapolis, Indiana. Caregivers were approached for enrollment in this qualitative study if their child was: (1) born prematurely and admitted to the NICU during their infancy; or (2) experienced AKI or another kidney complication (such as a slow to normalize serum creatinine), which put them at risk for future CKD. Caregivers were eligible for this study if their child was between ages 2 and 25 years old, if they agreed to participate in the web-based session, and if they had no diagnosed cognitive disabilities.

Recruitment was conducted by phone as well as in the outpatient pediatric nephrology clinic at Riley Hospital for Children, part of Indiana University Health, in Indianapolis, Indiana. Permission to approach the caregiver was obtained from the nephrologist of record to ensure the child did not have any medical treatments or conditions that could deter participation in the session. Informed consent was obtained from each study participant. Study participants were given a US $100 Amazon gift card for their engagement.

**Exploring and Co-Design**

We held 2 virtual sessions that were identical in purpose and methods but engaged different stakeholder groups. The first session included clinicians, and the second session included caregivers (Multimedia Appendix 1).

Specific activities included during the exploring and co-design sessions included:

**Empathy Mapping**

Empathy mapping is a generative method in which stakeholders are asked to intentionally speak about different aspects of an
experience (thinking and feeling, hearing, seeing, and saying and doing) [24]. Stakeholders (clinicians and caregivers) were asked to address each of these areas based on the following prompt: “After their child has received life-saving drugs in the NICU, parents are told that their child will need lifelong kidney monitoring. Help us understand this conversation.” To understand the context, stakeholders were also asked to describe where, when, and how this conversation took place. In addition, stakeholders were asked about the barriers to and drivers of lifelong kidney monitoring.

**Co-Design**

Co-design refers to the practice of guiding caregiver and clinician co-designers in the design development process [25]. The following co-design methods were used:

1. **Concept generator**: we created a worksheet in Miro to help caregiver and clinician co-designers diverge and converge on the function and form of a potential tool. It included the following instructions:

   *We need to develop a tool to help patients and their families overcome their barriers to long-term kidney monitoring. Let’s think creatively about what that tool could be.*

2. **Prototyping**: creating a rough version of a solution (a prototype) gave designers and caregiver and clinician co-designers the opportunity to make rough ideas tangible to quickly gain feedback and make iterations. Prototypes displayed the approximate of the solution or part of the solution [26]. How the prototype looked at this stage was less important than the conversation about why features were included and what problems each feature solved. We created a worksheet in Miro to help stakeholders create their prototypes.

   *Rose, Thorn, Bud*: Rose, Thorn, Bud was a reflective activity used during the session to get stakeholders to intentionally think about each prototype and provide feedback [27]. As a group, stakeholders focused on 1 prototype at a time and then shared 3 things: something that they thought was working well (a rose), something that presented a challenge (a thorn), and something that represented an opportunity or idea with potential (a bud).

**Analysis of Exploring and Co-Design Sessions**

Data (including the developed products, notes, and transcripts) from the sessions were analyzed using John Kolko’s methods of analysis and synthesis, using a creative process to connect research insights with design patterns to generate well-grounded design ideas [28]. These data were grouped by affinity or similarity of content, with each group given a heading to summarize its content. The resulting affinity diagrams spatially organized the data into groups based on similarity of content and represented the full picture of the data organized by theme [29]. Next, an analysis team created visual models of the themes and how they were interrelated [28].

Models included a refined empathy map, a communication opportunities map, and a grouping of “must have,” “can’t have,” and “nice to have” features for the communication tool. During model-building, a total of 2 “must have” and “nice to have” continuums were created. Each of the educational content and bonus feature items were placed on their respective continuums as determined by participants. Discussion points collected during the sessions were placed below their related item in the continuum.

**Prototype Development**

The research team then moved to prototype development, which looked at the outcomes of analysis (“what is”) to build solutions for the future (“what could be”) using the following synthesis methods:

**Brainstorming Potential Challenges to Solve**

To diverge further on what the solution could be, the team identified underlying challenges within the main objective. Asking “how might we...” allowed the research team to think beyond first instinct responses and use a divergent mindset to come up with many potential ideas for solutions. The research team then converged on the challenges that best fit the objective and what was learned from the analysis.

**Brainstorming Potential Solutions for Selected Challenges**

The research team asked one “how might we...” question at a time and listed as many solutions as they could. The research team used a divergent mindset, limited judgement, and focused on quantity over quality. Thinking broadly allowed for the generation of out-of-the-box solutions that could be examined for valuable elements that could be implemented into a final tool.

**SCAMPER Method to Diverge on Additional Solutions**

To further diverge, the research team used the SCAMPER method to create new solutions by manipulating already-stated solutions [30]:

- Substitute: what could you substitute or change?
- Combine: could two or more ideas or pieces be combined into something else?
- Adapt: what could be tweaked to improve the solution?
- Modify: could some solutions be changed to be improved?
- Put to another use: could solutions apply to another use?
- Eliminate: what could we take away from these solutions to improve them?
- Reverse: would rearranging elements improve solutions?

The research team used each of these prompts to create new solutions based on the existing process or solutions from the previous step. Following the divergent stage, the research team reviewed the list of solution ideas and voted for those they thought were the most appropriate and interesting.

**Prototyping**

The research team individually created prototypes of the tool inspired by the converged list of solution ideas, allowing the research team to explore additional ideas that could be included in the final tool. Refined prototypes were then used to get feedback from the stakeholders.
Prototype Refinement

We held 1 virtual session with a subgroup of clinicians and caregivers by Zoom to evaluate the prototypes developed using Miro.

1. **Sorting to Prioritize Prototype Content and Features:** Study participants were presented with a list of potential educational and informational elements identified as either “must have” or “nice to have” by the research team. Study participants were then asked to discuss and sort each of these into one of the two categories themselves. This same approach was taken with a list of bonus features the tool could include. This activity allowed for potential elements to be categorized based on the perspectives of the stakeholders, not just the research team.

2. **Prototype Feedback using Rose, Thorn, Bud:** Study participants were shown 2 prototypes. Each prototype had 3 main elements: information and education, bonus help, and appointment reminders. Prototype A focused primarily on digital solutions, while prototype B focused on analog solutions. The research team presented both prototypes to the study participants, asked for clarifying questions, then worked through the same Rose, Thorn, Bud activity used in phase 1 to get feedback for each prototype. This activity helped the research team understand elements of the prototypes that stakeholders liked and disliked.

3. **Frankenstein Prototypes:** With knowledge and opinions about what should go into the tool, study participants were asked to build new prototypes using their favorite elements from prototypes A and B. With the ability to mix, match, and create new elements, the research team could see what the participants prioritized.

Analysis of Prototype Refinement

We used affinity diagramming to group the feedback provided during Rose, Thorn, Bud. Through discussion within the research team, feedback from participants was arranged into groups and given thematic headings. These headings were used to identify key elements that study participants liked, did not like, and saw as having potential in the prototypes presented to them, allowing the research team to make final decisions about how to refine the prototypes. The research team then reviewed each item on the continuum and made decisions about what should be included in the final communication tool. Decisions were made based on feasibility and how well the item would address the original objective.

Ethics Approval

This study was approved by Indiana University’s institutional review board (protocol #11958), by whom it was deemed minimally risky.

Results

Participants

The exploring and co-design sessions included 15 participants (7 clinicians and 8 parents), while the prototype refinement session included 10 participants (6 clinicians and 4 caregivers). We approached 20 clinicians (7/20, 35% participation rate) and 32 caregivers (8/32, 25% participation rate; **Multimedia Appendix 1**). All the patients represented by caregivers in this study were discharged from the hospital and were currently seeing pediatric nephrology for monitoring of kidney health or management of CKD. See **Table 1** for demographic characteristics for the study participants in the exploring and co-design sessions.
Table 1. Demographic characteristics for study participants in the exploring and co-design sessions.

<table>
<thead>
<tr>
<th></th>
<th>Clinicians, N=7</th>
<th>Caregivers, N=8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4 (57)</td>
<td>6 (75)</td>
</tr>
<tr>
<td>Male</td>
<td>3 (43)</td>
<td>2 (25)</td>
</tr>
<tr>
<td>Age (in years), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-44</td>
<td>6 (86)</td>
<td>7 (88)</td>
</tr>
<tr>
<td>45-64</td>
<td>1 (14)</td>
<td>1 (12)</td>
</tr>
<tr>
<td>65 and older</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1 (14)</td>
<td>1 (12)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1 (14)</td>
<td>1 (12)</td>
</tr>
<tr>
<td>White</td>
<td>5 (72)</td>
<td>6 (75)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1 (14)</td>
<td>2 (25)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>6 (86)</td>
<td>6 (75)</td>
</tr>
<tr>
<td>Clinical subspecialty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General pediatrics</td>
<td>1 (14)</td>
<td>N/A</td>
</tr>
<tr>
<td>Neonatal and perinatal medicine</td>
<td>2 (28)</td>
<td>N/A*</td>
</tr>
<tr>
<td>Pediatric nephrology</td>
<td>4 (58)</td>
<td>N/A</td>
</tr>
<tr>
<td>Child’s current age (in years), mean (SD)</td>
<td>N/A</td>
<td>6 (4)</td>
</tr>
</tbody>
</table>

*N/A: not applicable.

Caregiver Experience

The caregiver experience began with their infant’s admission to the NICU. Sometimes, caregivers and clinicians expected that an infant would require immediate medical intervention after birth, while other times it was unexpected. Either way, infants required medical care in the NICU, with their caregiver as the primary decision maker. About this moment, one participant said (paraphrased): “I sat and looked at this perfect baby and they’re telling us she has all these challenges.” During the course of medical care, parents were often involved in difficult decisions or treatment decisions, such as clinicians recommending the use of life-saving medications and treatments that could harm their kidneys (eg, nephrotoxic medications, surgery, and other interventions; Multimedia Appendix 2).

In addition to making decisions critical to their infant’s care, the physical location of the decision placed additional stress on caregivers. Frequently, discussions between caregivers and clinicians occurred in the NICU, sometimes privately but often in the proximity of other patients and passersby. Caregivers described an overwhelming scene with many new sights and sounds, hopes and fears, high and low emotional points, and advanced levels of stress and fatigue. As doctors presented the treatment options and the implications of those options, caregivers found it easy to lose focus and not remember all the details of the conversation. They may or may not remember being informed that the child would need lifelong kidney monitoring due to potential kidney damage from life-saving treatments (Multimedia Appendix 2).

Barriers and Drivers to Monitoring

Some caregivers recalled that clinicians suggested the need for kidney monitoring at the time of discharge. Both caregivers and clinicians readily acknowledged barriers to and drivers of long-term kidney monitoring. Caregivers shared that many of their pediatricians and other health care clinicians agreed with or reinforced the need to monitor the patient’s kidneys; however, at least one caregiver was told that it was not necessary by their pediatrician. Adherence to kidney health monitoring, in addition to other treatments that may be required following their NICU admission, posed more immediate challenges, such as the difficulty of their young child tolerating a blood draw or urine collection. Caregivers weighed these barriers versus the drivers of early identification of kidney problems, improved care for their child, and saving money over time. Barriers to and drivers of long-term kidney monitoring are summarized in Figure 1 and Multimedia Appendix 2.
Communication Considerations and Opportunities

Stakeholders reported positive and negative aspects of the communication of medical information, both generally and about the implications of kidney injury specifically (Multimedia Appendix 2). Caregivers noted that due to the stressors experienced by caregivers and the challenges of learning and memory retention in the NICU, clinicians should offer information about kidney health and long-term kidney monitoring at multiple points throughout the NICU admission, including at the time of administering medications or therapies that may contribute to kidney injury, at discharge as part of the discussion of follow-up care needed, and at follow-up appointments. Figure 2 shows a model of caregiver experience with communication opportunities identified.

Figure 1. Barriers and drivers to long-term kidney monitoring for children after the neonatal intensive care unit (NICU).
Contents and Features
Study participants sorted educational content and potential features into “must have” and “nice to have” categories. Each of the educational contents was placed on their respective continuums based on where it was placed by study participants (Table 2). For example, “questions to ask clinicians” was placed in the “must have” section of the education and information continuum because both groups sorted it as a “must have,” while “NICU guide” was placed both in the “must have” section and the “nice to have” section because 1 group sorted it as a “must have” item and the other as a “nice to have item.”
The research team then reviewed each item on the continuum and made decisions about what should be included in the final tool. The research team also discussed which of the items from the middle 2 sections should be included. The research team reviewed the bonus features and decided which of these to include in the final tool. Decisions were made based on how well the item would address the original objective.

**Prototypes and Feedback**

The research team created 2 prototype web pages to illustrate what a final tool might look like (Multimedia Appendix 3). For example, the home page included information about poor kidney development, potential kidney injury in the NICU, and how this may lead to the need for long-term kidney monitoring. It also contained a still from a video that might exist where a clinician explains NICU kidney injuries. The home page acts as the basic information for caregivers, while the rest of the site offers additional details. The menu items included: “about kidney monitoring,” “common kidney tests,” “talking with your child’s doctor,” and “caregiver support.” Caregivers and clinicians reviewing the prototypes were supportive of the categories of information and content provided. They also appreciated the overall design of the prototype webpage. In general, they wanted caregiver stories with diverse people and languages, as well as more detailed information and research.

**Discussion**

We conducted a qualitative study examining caregiver attitudes and preferences toward the communication of kidney health by clinicians in the NICU setting. Our results suggest opportunities for improving communication about the risk of long-term kidney disease between caregivers and clinicians. Caregivers’ primary concerns were the type and depth of information conveyed and...
the time at which it was communicated. Both caregivers and clinicians emphasized the importance of collaboration between the NICU team and the primary care provider to ensure they were on the same page about the necessity of kidney monitoring.

This study represents the first attempt, to our knowledge, to develop a set of clear approaches to communicating kidney health and the risk of CKD in the NICU. Our findings are in concert with a recently published survey of caregivers with infants diagnosed with necrotizing enterocolitis during their NICU admission [17]. Both studies found that caregivers desired accurate and timely information to inform care and improve communication. Furthermore, other studies suggest that information gathering is an important coping mechanism for stress while their child is in the NICU [31]. Education-based programs have additional benefits for caregivers, including improved parental mental health outcomes, stronger beliefs in their parental role, and increased parental engagement [32]. The timely and family-centered provision of information and education is an essential aspect of family-centered care, which has increased parent engagement and satisfaction as it has become more widely used in NICUs over the last decade [33].

One challenge in neonatal kidney health clinical care and research studies is the low rate of kidney-specific follow-up for infants [34,35]. Studies suggest that, while multifactorial, contributing factors include poor provider and caregiver awareness of the risk of long-term kidney disease, a lack of family communication, and a perceived inability to change the course of disease with care [36]. Furthermore, siloing of care and electronic health care records which do not follow patients from health encounter to health encounter limit the ability of caregivers and clinicians to carry health information with them throughout the medical system. The development of improved communication with caregivers during and after their NICU stay is paramount to improving not only clinical care but also research studies of long-term kidney health, which are often stymied by poor retention. Our approach to kidney health communication was developed not by expert consensus of clinicians, as is often the case in similar studies, but by directly engaging with caregivers who have had infants admitted to the NICU who are at risk of long-term kidney disease. We believe this will result in a far more effective communication strategy that is more acceptable to families and increases the efficacy of subsequent follow-up.

There are several important limitations to this study. First, owing to the relatively small sample size and narrowness of the study population (eg, caregivers of infants at risk for CKD in the NICU), it is difficult to ascertain the broad generalizability of these findings. However, we attempted to recruit caregivers of various ages and backgrounds, at varying time periods out from their child’s NICU stay (eg, 6 months post-NICU discharge vs 2 years post-NICU discharge) in order to improve generalizability to our larger population. Second, the design methods used are novel in health-related research, but they have been well-established in service and product design. Finally, the subjects we recruited were a convenience sample of nonconsecutive caregivers seen at our pediatric nephrology clinic who were willing to participate in research and may not accurately represent a random sample of our patient population.

Despite these limitations, this study represents an important first step in improving communication about kidney health to caregivers and families of those at high risk of kidney disease. The next step in this project is to further develop this communication tool based on caregiver and clinician guidance and to implement the tool in the NICU. Based on the above results and guidance from participants in this project, we are developing a website for family-centered kidney health information and plan to continue to gather input from caregivers to better understand the best ways to present and organize information, how to provide real-world experience and perspectives, and what information caregivers want at specific times during and after their child’s NICU admission. Caregivers of infants admitted to the NICU will be given access to the revised communication tool developed in this study. We will then further assess the impact of the communication tool on their understanding of kidney health, the risk of long-term kidney disease, and follow-up patterns.

Acknowledgments

We thank all of the clinicians and caregivers who participated in this work. We thank the Agency for Healthcare Research and Quality for their generous support for this study through the Infrastructure Development Program in Patient-Centered Outcomes Research R24HS022434 (WEB Jr) and Research Jam. Research Jam is supported by the Indiana Clinical and Translational Sciences Institute, funded in part by the National Institute for Advancing Translational Science TR001107 (WEB Jr).

The author MS was supported in part by the Physician Scientist Fund at IUSM. The funding sources for this study had no role in the design and conduct of the study, collection, management, analysis, and interpretation of the data, preparation, review, or approval of the manuscript, and decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Process of human-centered design process to create family-centered communication tool.

[ PNG File, 364 KB - jopm_v151e45316_app1.png ]
Multimedia Appendix 2
Parent comment on barriers and drivers of kidney monitoring and communication strategies.

[DOCX File, 15 KB - jopm_v15i1e45316_app2.docx]

Multimedia Appendix 3
Preliminary prototypes developed.

[PNG File, 495 KB - jopm_v15i1e45316_app3.png]

References


Examining Patient Engagement in Chatbot Development Approaches for Healthy Lifestyle and Mental Wellness Interventions: Scoping Review

Chikku Sadasivan¹, BSc; Christofer Cruz¹, BSc, RD; Naomi Dolgoy², MOT, CLT, PhD; Ashley Hyde¹, RN, PhD; Sandra Campbell¹, MLS; Margaret McNeely², PT, PhD; Eleni Stroulia³, PhD; Puneeta Tandon¹, MSc, MD

¹Department of Medicine, University of Alberta, Edmonton, AB, Canada
²Department of Physical Therapy, University of Alberta, Edmonton, AB, Canada
³John W Scott Health Sciences Library, University of Alberta, Edmonton, AB, Canada
⁴Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Puneeta Tandon, MSc, MD
Department of Medicine
University of Alberta
130 University Campus
Edmonton, AB, T6G 2R7
Canada
Phone: 1 780 492 9844
Email: ptandon@ualberta.ca

Abstract

Background: Chatbots are growing in popularity as they offer a range of potential benefits to end users and service providers.

Objective: Our scoping review aimed to explore studies that used 2-way chatbots to support healthy eating, physical activity, and mental wellness interventions. Our objectives were to report the nontechnical (eg, unrelated to software development) approaches for chatbot development and to examine the level of patient engagement in these reported approaches.

Methods: Our team conducted a scoping review following the framework proposed by Arksey and O’Malley. Nine electronic databases were searched in July 2022. Studies were selected based on our inclusion and exclusion criteria. Data were then extracted and patient involvement was assessed.

Results: 16 studies were included in this review. We report several approaches to chatbot development, assess patient involvement where possible, and reveal the limited detail available on reporting of patient involvement in the chatbot implementation process. The reported approaches for development included: collaboration with knowledge experts, co-design workshops, patient interviews, prototype testing, the Wizard of Oz (WoZ) procedure, and literature review. Reporting of patient involvement in development was limited; only 3 of the 16 included studies contained sufficient information to evaluate patient engagement using the Guidance for Reporting Involvement of Patients and Public (GRIPP2).

Conclusions: The approaches reported in this review and the identified limitations can guide the inclusion of patient engagement and the improved documentation of engagement in the chatbot development process for future health care research. Given the importance of end user involvement in chatbot development, we hope that future research will more systematically report on chatbot development and more consistently and actively engage patients in the codevelopment process.

(J Particip Med 2023;15:e45772) doi:10.2196/45772

KEYWORDS
chatbots; virtual assistants; patient involvement; patient engagement; codevelopment

Introduction

Growing evidence supports the use of digital technology in healthy eating, physical activity, and mental wellness interventions. Several systematic reviews on these digital health interventions (DHIs) have identified their promise in managing chronic diseases [1-6]. Specifically, DHIs have proven impacts on reducing risk factors for chronic diseases [3,4] by increasing...
physical activity, reducing body mass index [6], and improving patient psychosocial well-being [2]. Further, DHIs can help overcome barriers to access to mental health support for individuals with chronic conditions [1]. Although these DHIs are useful in vulnerable chronic disease populations [5,7], they face challenges, including limited user adoption, low engagement, and high attrition rates [8-11].

Chatbots are artificial intelligence (AI) programs that converse with humans through natural language in text or speech [12]. There is a growing body of evidence that the integration of chatbots into DHIs may provide support [13-17] by increasing patient engagement [13], intervention adherence [13], and the acceptability and efficacy of lifestyle and wellness interventions [15-17]. Additionally, chatbots offer a range of potential benefits to end users and service providers, most notably allowing for more scalable, cost-efficient, and interactive solutions [12]. Although developments in AI and computer science have improved the ability of chatbots to mimic human agents, the acquisition of a relevant data set with which to train chatbots remains challenging. User-centered design with public and patient involvement (PPI) may offer a potential solution [18-20]. By engaging key stakeholders, PPI can help produce better-quality interventions relevant to end users’ needs [18], resulting in benefits such as increasing intervention acceptability, effectiveness, and sustainability [19]. Drawing on evidence across other digital health care innovations, the proposed benefits of PPI fundamentally include the development of interventions that are both usable by and relevant to patients [19]. Recognizing the limited data available to guide the role of PPI in digital health innovation, experts have called for the meaningful involvement of patients from the beginning of the development process to allow for the co-creation of relevant, valuable, and acceptable digital health solutions [20].

This scoping review aimed to map the literature on studies using chatbots to engage in 2-way natural language interaction (voice- or text-based input) to aid the delivery of healthy eating, physical activity, and mental wellness interventions. The specific objectives of this review were: (1) to report the nontechnical (eg, unrelated to software development) approaches for chatbot creation and (2) to examine the level of patient engagement in these reported approaches. Although the technical software development steps are essential to creating chatbots, this review focused on the nontechnical approaches for chatbot development as these are less explored and more likely to involve patient participation. To our knowledge, this is the first scoping review to systematically explore these objectives.

**Methods**

**Study Design**

This scoping review was conducted using the framework proposed by Arksey and O’Malley [21] and later refined by Levac et al [22]. The Arksey and O’Malley framework consists of the following five steps: (1) identify a research question, (2) identify relevant studies, (3) select studies, (4) chart the data, and (5) summarize and report the results [21]. Two research questions guided the review:

1. Outside of the technical software development processes, what approaches are described for the development of chatbots that support healthy eating, physical activity, and mental wellness interventions?
2. What is the extent of patient engagement in these approaches?

**Study Team**

Our multidisciplinary study team included 2 graduate student researchers (CS and CC), a health sciences librarian (SC), 2 postdoctoral fellows with backgrounds in clinical care and scoping reviews (ND and AH), a professor of medicine (PT), a professor of physiotherapy (MM), and a professor of computing science (ES).

**Search Strategy**

A health sciences research librarian (SC) was consulted to develop a search strategy that used concepts from our research questions. The search strategy (Textbox 1) included a combination of subject headings and keywords, including health, chatbots, and lifestyle or wellness components. Searches were adjusted appropriately for each database. Nine electronic databases were searched in July 2022 including OVID MEDLINE, OVID Embase, OVID PsycINFO, EBSCO CINAHL, Scopus, IEEE Explore, Proquest Dissertations and Theses Full Text, Cochrane Library, and PROSPERO (International Prospective Register of Systematic Reviews). No publication date limit was applied to the search, as the literature on chatbots and virtual conversation agents is naturally self-limiting. After conducting the search, the results were imported into Covidence systematic review management software and duplicates were removed [23]. Covidence is a “web-based collaboration software platform that streamlines the production of systematic and other literature reviews” [23]. The full text of the search strategy is in Multimedia Appendix 1.
Textbox 1. Search strategy used for OVID PsycINFO database.

# Searches
1. (chatbot* or “im bot” or “im bots” or “instant message bot*” or “conversational agent*” or “virtual agent*”).mp.
2. “Diets”/
3. “Health Promotion”/
4. “Intervention”/
5. “Physical Activity”/
6. “Nutrition”!
7. “Weight Loss”!
8. “Sedentary Behavior”/
9. (lifestyle* or health* or medic* or nursing or nurse* or disabilit* or elder* or “senior citizen*” or patient* or exercise or “physical activit*” or motivational).mp.
10. 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
11. 1 and 10

Eligibility Criteria
Included publications were those written in English and published in peer-reviewed journals. Included studies all had an intervention supporting healthy eating, physical activity, and mental well-being. All studies required a chatbot that communicated with users through a 2-way natural language interaction. Inclusion criteria for participants consisted of adolescents (age >10 years old) as defined by the World Health Organization [24] or adult populations. Studies were excluded if they involved additional technologies or chatbot service delivery beyond the scope of this review (ie, embodied conversation agents, humanoid and social robots, wearable technology, Internet of Things (IoT), virtual avatars, interactive voice assistants, or chatbots delivering therapy to clients). Studies were also excluded if they only described an intervention but did not conduct or test one. Chatbots designed to replace a therapist’s role were excluded, as were papers that did not present original results (ie, reviews and protocol papers). Randomized controlled trials (RCTs) were included in recognition that they often contain valuable insights into the development process, particularly when the authors did not publish a formative manuscript.

Data Extraction
One reviewer (CS) extracted the data from included articles using a standardized Microsoft Excel form. General and specific data were extracted, including author, publication year, journal, study setting, study design, sample size, participant demographics (age, sex, and chronic disease where applicable), intervention type, chatbot type, chatbot development approaches, and assessment of patient involvement in development.

Patient involvement was assessed using the Guidance for Reporting Involvement of Patients and Public (GRIPP2) short-form checklist [26]. The GRIPP2 checklist was applicable for our objectives as it was designed to enhance the quality of patient and public involvement (PPI) reporting in health technology assessment and health research [26], and because it could be used retrospectively to measure the quality of PPI reporting in publications and reports [27]. Table 1 depicts the GRIPP2 checklist as we used it to assess PPI in chatbot development. The GRIPP2 awards points across 5 items that describe public engagement and involvement.
Table 1. How the Guidance for Reporting Involvement of Patients and Public (GRIPP2) reporting checklist was used to grade patient and public involvement in chatbot nontechnical development.a

<table>
<thead>
<tr>
<th>Section and topic</th>
<th>Specifics for engagement in chatbot-related development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Aim</td>
<td>Report the aim of PPIb in chatbot development</td>
</tr>
<tr>
<td>2. Methods</td>
<td>Provide a clear description of the methods used for PPI in chatbot development</td>
</tr>
<tr>
<td>3. Study results</td>
<td>Outcomes: Report the results of PPI in chatbot development, including both positive and negative outcomes</td>
</tr>
<tr>
<td>4. Discussion and conclusions</td>
<td>Outcomes: Comment on the extent to which PPI influenced chatbot development overall. Describe positive and negative effects</td>
</tr>
<tr>
<td>5. Reflections or critical perspective</td>
<td>Comment critically on chatbot development, reflecting on the things that went well and those that did not, so others can learn from this experience</td>
</tr>
</tbody>
</table>

aAdapted from Staniszewska et al [27].
bPPI: patient and public involvement.

Results

Search Results

Figure 1 shows the search results; 3089 publications were retrieved from the database searches, and 882 duplicates were removed, leaving 2207 studies to screen. At the title and abstract screening stage, there was “fair” agreement between reviewers (Cohen κ=0.309, proportionate agreement=0.967). After completing the title and abstract screening, 2140 publications were removed as they did not meet the inclusion criteria. Reading the full text of the remaining 67 publications resulted in a further 51 publications being excluded, with the exclusion reasons documented in Figure 1. At the full-text review stage, there was “almost perfect” agreement (Cohen κ=0.843, proportionate agreement=0.941). In total, 16 publications were included in this review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram of included and excluded studies.

Description of Included Studies

Table 2 shows the description of the included studies and their chatbot interventions. The included studies were conducted in 4 countries, with 50% (8/16) of the studies conducted in Canada [28-35]. Six studies were conducted in Switzerland [36-41], 1 study was conducted in Saudi Arabia [42], and 1 study was conducted in Korea [43]. The majority of the studies (14/16) were conducted in a health care setting [28-40,43], with the remaining 2 studies in a computing science setting [41,42]. All but one of the included studies [31] were published in 2020 or later.
Table 2. Descriptive summary of included studies, chatbots, and their development.

<table>
<thead>
<tr>
<th>Study and country</th>
<th>Study type</th>
<th>Chatbot intervention</th>
<th>Approaches for development</th>
<th>Identified development approaches</th>
<th>Patient engagement (GRIPP2&lt;sup&gt;a&lt;/sup&gt;)</th>
</tr>
</thead>
</table>
| Alghamdi et al [42], Saudi Arabia | Randomized controlled trial | Text-based nutrition chatbot for patients with celiac disease | • Literature review of existing health behavior change models. Investigated the pros and cons of each model to guide development of a health behavior change model to structure the chatbot's content  
• Interviews with expert users (from patient population diagnosed with celiac disease 4+ years ago, patient’s parent, dietitian supervising patient for 4+ years, gastroenterologist treating celiac disease patient for 4+ years)  
• Questionnaires for patients with celiac disease to understand symptoms and technology use preferences | • Literature review  
• Patient interviews  
• Collaboration with knowledge experts | Unable to assess |
| Davis et al [36], Switzerland | Nonrandomized experimental study | Text-based exercise and nutrition chatbot | • Development outsourced to a software company; did not report any steps taken for development | • None identified | Unable to assess |
| Dhinagaran et al [28], Canada | Feasibility study | Text-based exercise, nutrition, and wellness chatbot for patients with diabetes | • Needs assessment conducted in an earlier publication  
• Literature review of systematic reviews and clinical guidelines for evidence-based content development to develop content  
• After a 4-week pilot feasibility study, conducted follow-up interviews to understand patient views of the chatbot and to gain ideas for improvement | • Literature review  
• Patient interviews | Unable to assess |
| Figueroa et al [37], Switzerland | User design study | Text-based exercise chatbot | • Qualitative interviews during prototype testing to assess opinions and knowledge of chatbots as personal health coaches, technology use, digital literacy, and privacy considerations of chatbots in general  
• Wizard of Oz procedure. Participants completed a 20-minute SMS text messaging conversation with a simulated chatbot  
• Chatbot prototype testing. Participants texted the prototype for 10-20 minutes. Directly after the testing period, participants had a semistructured interview via videoconference regarding the chatbot’s ease of use, usefulness, humanness, and sustainability  
• Co-design workshop for participants to take part in development of ideas for chatbot use and design. These workshops were held over Zoom and ideas were visualized on Google Jamboard | • Patient interviews  
• Wizard of Oz procedure  
• Prototype testing  
• Co-design workshops | Met criteria on GRIPP2 checklist points 2, 4, and 5. Provided a clear description of the methods used for PPI<sup>b</sup>, commented on how PPI influenced the study, and on successful and unsuccessful aspects of the study relating to PPI |
| Gabrielli et al [29], Canada | Proof-of-concept study, mixed methods | Text-based wellness chatbot | • Intervention design. The intervention, targets, and components were defined to specify clinically relevant effects on users and to refine the intervention components. This was done by a team of 3 clinical psychologists, 2 users, and behavior change experts  
• Preliminary testing. A proof-of-concept implementation of the digital intervention and chatbot to examine engagement and effectiveness with a convenience sample of university students | • Collaboration with knowledge experts  
• Prototype testing | Unable to assess |
<table>
<thead>
<tr>
<th>Study and country</th>
<th>Study type</th>
<th>Chatbot intervention</th>
<th>Approaches for development</th>
<th>Identified development approaches</th>
<th>Patient engagement (GRIPP2)</th>
</tr>
</thead>
</table>
| Gabrielli et al [30], Canada | Pilot, co-design study | Text-based wellness chatbot | • Co-design workshop. The students used and commented on a prototyped session of the chatbot intervention to collect their needs and preferences on the following: the chatbot’s look and feel, the type of content and duration of the session, their unmet expectations regarding the prototype, and suggested improvements  
• Feasibility test. This formative study aimed to assess the perceived value of the coaching intervention and to check the user experience with intervention to refine content | • Co-design workshops | Met criteria on GRIPP2 checklist point 2. Provided a clear description of the methods used for PPI |
| Greer et al [31], Canada | Randomized controlled trial | Text-based wellness chatbot for patients with cancer | • Literature review of the Stress and Coping theory and the Broden-and-Build theory of positive emotion and focused on the teaching and practice of 8 positive psychological skills. Created lessons based on this review for the chatbot to deliver  
• Interviews and focus groups as formative work to refine content for the chatbot format and inform adaptation for delivery to a young user base with a shared experience of cancer treatment | • Literature review  
• Patient interviews | Unable to assess |
| Issom et al [38], Switzerland | Usability study | Text-based exercise, nutrition, and wellness chatbot for patients with SCD | • Literature review of evidence-based knowledge of SCD self-management, in addition to consulting the World Health Organization’s handbooks on how to implement text-based mHealth interventions to help with dialogue design | • Literature review | Unable to assess |
| Krishnakumar et al [32], Canada | Nonrandomized experimental study | Text-based exercise and nutrition chatbot for patients with type 2 diabetes mellitus | • Literature review to develop a lesson plan of the program. This was based on the American Association of Diabetes Educators’s AADE7 self-care behaviors | • Literature review | Unable to assess |
| Larbi et al [39], Switzerland | Usability study | Text-based exercise chatbot | • Literature review of behavior change interventions Summarized and briefly reported 4 steps in development: strategy planning, design, implementation, and testing. As part of strategy planning, psychology and public health experts were interviewed  
• Also stated that the development of the prototype involved 3 steps: requirement analysis, concept development, and implementation. Reporting did not go into any further detail | • Literature review | Unable to assess |
| Maenhout et al [40], Switzerland | Development pilot study | Text-based exercise, nutrition, and wellness chatbot | • Intervention planning through a scoping review of literature, conducting focus groups, and consulting web-based chat threads for a youth helpline. Focus groups addressed: content preferences, design preferences, questions that the chatbot would be asked, and answers that were expected from the chatbot  
• Intervention optimization through conducting a log data analysis during pretesting. A prototype of the chatbot was developed and pretested by the target users. The prototype was developed based upon guidance from phase 1 focus groups. Conversation logs were closely monitored to refine and fine-tune the chatbot. A question list was formed at the end of this prototype testing phase, 37 new (and practical) questions originated that were not covered in the chat threads and focus groups | • Literature review  
• Patient interviews  
• Prototype testing | Met criteria on GRIPP2 checklist point 2 |
<table>
<thead>
<tr>
<th>Study and country</th>
<th>Study type</th>
<th>Chatbot intervention</th>
<th>Approaches for development</th>
<th>Identified development approaches</th>
<th>Patient engagement (GRIPP2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maher et al [33], Canada</td>
<td>Proof-of-concept study</td>
<td>Text-based exercise and nutrition chatbot</td>
<td>• Did not report how the chatbot was developed; the methods section described how the pilot study was conducted</td>
<td>None identified</td>
<td>Unable to assess</td>
</tr>
<tr>
<td>Pecune et al [41], Switzerland</td>
<td>Nonrandomized experimental study</td>
<td>Text-based nutrition chatbot</td>
<td>• Literature review of persuasive systems, recommender systems, and food-related experiments. Collected a food database by regrouping the 40 ingredients that people most frequently cook and eat for dinner. These data were collected from hundreds of participants through questionnaires. • Completed a pilot study to determine what the critical elements are for recipe recommendation systems. Also, completed this quasi-experimental study to understand the efficacy of different chatbot characteristics with the target end user group</td>
<td>Literature review</td>
<td>Unable to assess</td>
</tr>
<tr>
<td>Piao et al [43], Korea</td>
<td>Usability study</td>
<td>Text-based exercise chatbot</td>
<td>• Needs assessment through web-based surveys to assess daily routines of office workers (the target group). This was used to determine daily activities that were measurable and easy to execute. These became a part of the goal setting in the intervention. • Chatbot design was guided through a review of the literature and to determine a theoretical model for the chatbot’s basis: the habit formation model. • Conducted this formative usability test prior to the randomized controlled trials below to identify issues and make revisions</td>
<td>Literature review • Prototype testing</td>
<td>Unable to assess</td>
</tr>
<tr>
<td>Piao et al [35], Canada</td>
<td>Randomized controlled trial</td>
<td>Text-based exercise chatbot</td>
<td>• Literature review of extrinsic and intrinsic reward systems. • Steps for development were documented in the usability study described above</td>
<td>Literature review</td>
<td>Unable to assess</td>
</tr>
<tr>
<td>To et al [34], Canada</td>
<td>Nonrandomized experimental study</td>
<td>Text-based exercise chatbot</td>
<td>• Development was outsourced for technical development by SmartAI. Did not report if the research team was involved in any other steps for development</td>
<td>None identified</td>
<td>Unable to assess</td>
</tr>
</tbody>
</table>

aGRIPP2: Guidance for Reporting Involvement of Patients and Public.
bPPI: patient and public involvement.
cSCD: sickle cell disease.

**Study Design and Interventions**

Three of the included studies were RCTs [31,35,42], 4 were nonrandomized experimental studies [32,34,36,41], 3 were user-design and development studies [30,37,40], 3 were usability studies [38,39,43], 1 was a feasibility study [28], and 2 were proof-of-concept studies [29,33].

Fifteen of the 16 included studies reported the sample size; sample sizes ranged from 18 to 116 participants [34,37]. Participants’ age ranged from 12 to 69 years, with most participants being younger than 50 years old. When a specific chronic disease group was described, populations included patients with celiac disease [42], diabetes [28,32], cancer [31], and sickle cell disease [38]. Where reported, the inclusion of female participants ranged from 31.4% to 100% [37]. Five studies involved an exercise intervention [34,35,37,39,43]. Three studies included a mental wellness intervention for healthy coping, life skill coaching, and positive psychology skill building [29-31]. Two studies evaluated a nutrition intervention [41,42]. The remaining interventions combined exercise, nutrition, and mental wellness components [28,32,33,36,38,40]. Across all reviewed articles, the chatbots communicated with users through text.

**Study Findings**

There were several approaches used to guide the development and training of chatbots. In 3 of the included studies, the nonsoftware development approaches for chatbot development were not documented; therefore, no approaches were identified [33,34,36]. Thirteen studies reported approaches taken for chatbot development, with most studies reporting multiple approaches [28-32,35,37-43]. In 4 of the 13 studies, patients were engaged as knowledge experts or participants in co-design workshops [29,30,37,42]. In 6 of the 13 studies, patients were involved in the study as research participants and, as part of the study outcomes, were invited to share their views through
Collaboration With Patient and Clinician Partners as Knowledge Experts

During the early stages of chatbot planning, 2 studies consulted experts for chatbot development [29,42]. In both studies, patient partners were recognized as knowledge experts and included as part of the research team [29,42]. In the study with a nutrition chatbot for a celiac disease patient group, patients were recognized as experts alongside health care professionals, including dietitians and gastroenterologists [42]. In the mental wellness study, a team of 3 clinical psychologists took part in chatbot intervention development and content refinement alongside 2 users and a group of behavior change experts; this iterative process was used to adapt the chatbot’s intervention program, and audiovisual content to user needs through a clinical lens [29].

Co-design Workshops

Two studies used co-design workshops to allow patients to creatively engage in the development of content ideas, chatbot design, chatbot style elements, and chatbot use [30,37]. One study invited participants to collaborate and develop ideas together with the research team over Zoom (a web-based communication platform; Zoom Video Communications, Inc) by visualizing ideas on Google Jamboard software (a web-based whiteboard for idea sharing) [37]. Another study invited patients to use a prototyped session with the chatbot to collect their needs, content preferences, stylistic ideas, and suggestions for improvements [30].

Interviews With Patients

In 5 studies, patient interviews were conducted beforehand to guide chatbot development by exploring patient needs, perceptions, and experiences with chatbot use and healthy living [28,31,37,40,42]. In 1 study, interviews were administered during prototype testing and analyzed qualitatively [37]. Another study conducted this formative work through focus groups and interviews to collect information from young adults treated for cancer, the target end user population [31]. This information was then used to guide chatbot content development within a patient-centered lens. Follow-up interviews were conducted after interventions or chatbot exposure [28,40]. Questionnaires and surveys were also used in addition to interviews to collect similar information from patients [28,42].

Prototype Testing

Many included studies were nonexperimental or pilot studies used to assess the feasibility and measure usability. These formative studies can be considered a step for development before releasing and testing a mature chatbot in an RCT. For example, 1 study using a chatbot for an exercise intervention organized a 3-week formative usability study [43] to identify issues and make revisions before conducting an RCT [35].

WoZ Procedure

One study used the WoZ procedure [37] (where the technology is controlled by a human interface in chatbot development) as a step in their chatbot development. This procedure is administered by engaging participants in a 20-minute conversation with a simulated chatbot that was not automated but controlled manually by a researcher answering questions on the back end [37]. This step was developed to understand how the chatbot should interact with humans in a natural setting and to collect content-related information directly from participants [37].

Use of Existing Literature to Gain Evidence-Based Knowledge for Development

In 10 studies, initial literature reviews were completed to gain evidence-based knowledge to guide chatbot development [28,31,32,35,38-43]. In 3 of these 10 studies, a literature review was used to develop content from evidence-based sources, including self-management practices, clinical guidelines, and systematic reviews [28,32,38]. A mental wellness study incorporated this step into development by reviewing the psychological theories and practices used to create the lessons the chatbot would deliver [31]. In another study, a literature review of the existing health behavior change models was conducted to understand the pros and cons of each model, and to guide the development of a novel behavior change model to structure the chatbot’s content [42]. In 1 study, gray literature was sourced through web-based chat threads for a youth helpline, so researchers could better understand content topic preferences and expected answers [40]. Finally, 2 of these 10 studies reviewed the literature to learn more about reward systems and to identify a theoretical basis for chatbot development [35,43].

Patient Engagement and Public Involvement

Overall, the reporting of patient engagement in our included studies was limited making an assessment of PPI using the GRIPP2 challenging. Though 8 studies in our review reported involving patients, 5 provided inadequate detail, making assessing patient involvement impossible [28,29,31,42,43]. Specifically, these studies did not report on the aim of PPI, did not clearly articulate their methods, or did not discuss the role of PPI in their outcomes. The remaining 8 studies were not...
evaluated using the GRIPP2 because they did not report development approaches at all [33,34,36] or did not involve patients in the reported approaches [32,35,38,39,41].

Of the 3 studies we assessed using the GRIPP2, 1 study scored 3 points on the GRIPP2 Field [37], with the other 2 scoring 1 point [30,40]. Figueroa et al.’s study scored 3/5 on the GRIPP2 scale [37]. This study provided a clear description of the methods used for PPI, commenting on how PPI influenced the study and on successful and unsuccessful aspects of the study relating to PPI [37]. This study was also the only one that described 4 different approaches used for development, including co-design workshops, interviews, WoZ, and prototype testing. The authors noted that their co-design sessions “brought unexpected participant preferences and wishes, which were useful in developing subsequent versions” of their chatbot [37]. Further, they recognized the importance of engaging patients in design, testing, and dissemination to develop chatbot interventions that participants would use and benefit from. The remaining 2 studies, 1 by Gabrielli et al [30] and the other by Maenhout et al [40], were each awarded a single point on the GRIPP2 for clearly describing the methods used for PPI. The reporting was such that future researchers could replicate similar development approaches to actively engage patients in research design.

**Discussion**

**Principal Findings**

In this review, we described the non-technical approaches taken for chatbot development and evaluated the extent of patient engagement using the GRIPP2. While promising approaches were shared about the non-technical steps associated with chatbot development, the level of detail provided was often low, including how patients were involved in the process.

The limited level of detail speaks to the need to prioritize frameworks for implementing digital health tools [44,45]. This will involve a focus on increased formative, development, and feasibility studies and a shift to implementation research that considers embedding and sustaining interventions in context [44,45]. A more detailed focus on the developmental stages and implementation process in research would allow increased replicability of developmental approaches that actively engage patients and progress the field of chatbot research from the end user perspective. An example of this focus on the implementation process includes the formative work conducted by Islam and Chaudhry [46] while developing a chatbot to support the health care needs of patients during the recent COVID-19 pandemic. Their work is an example of detailed documentation of a replicable multi-phased chatbot design study, offering guidance for future research in this area [46]. Additional focus on implementation will ensure the production and monitoring of chatbots that provide quality care and service to patients across short- and long-term timelines [44]. This strategic planning also holds promise to better respond to the requirements of diverse user cohorts, especially those with lower levels of digital health literacy [47].

Although an attempt was made to evaluate the extent of the patient engagement process by the GRIPP2 patient engagement checklist, due to limited detail of reporting, this was only possible in 3 studies [30,37,40]. Many digital health solutions are plagued with low uptake and poor usability as they were developed with minimal patient involvement [48]. As user-centered design and patient engagement are known to improve the quality of research, engagement approaches throughout the research continuum could result in the identification of system requirements that would be otherwise missed, as well as result in a better understanding of patient needs, higher intervention engagement, and increased intervention effectiveness [49]. Some of the approaches we have identified in this review, including co-design workshops, the WoZ approach, patient interviews, and iterative prototype testing, represent ways researchers can creatively and actively engage patients throughout the development process. Co-design workshops foster a richer understanding of what patients “know, feel, and even dream” [50]. The WoZ approach is a widely accepted evaluation and prototyping methodology for developing human-computer interaction technology [51]. Engaging patients in iterative prototyping and user testing cycles has proven to improve the ease of use and adoption of these interventions [52]. In alignment with the literature, we recommend that researchers taking on health chatbot development projects consider adopting approaches such as co-design workshops, interviews, WoZ, and prototype testing.

Despite the available evidence supporting the benefit of patient involvement in intervention development, there are reasons why approaches that do not directly or actively involve patients may be more appealing to researchers. This notably includes challenges associated with recruitment, particularly when trying to avoid recruitment bias, and the time and resource intensity associated with the overall process [20]. The scarcity of patient involvement may also be related to an underappreciation of the potential benefits of patient involvement in digital health research and a limited understanding of how best to get patients involved [20]. Researchers and practitioners should be aware that there are many different approaches, strategies, and models to engaging patients in chatbot development. We have summarized some approaches in this review, and resources such as the Strategy for Patient-Oriented Research framework and the patient engagement in research plan offer practical information to guide patient involvement in the development process [53,54]. Patients can participate at all stages, helping to define health care problems, identify solutions, participate as co-designers of an intervention, and refine the evaluation process [19]. Figure 2 offers the direction in informing future research in patient-oriented chatbot development for lifestyle and wellness interventions, including the application of multifaceted means of patient engagement, use, and thorough documentation of approaches to enhance chatbot development, and clear and replicable reporting of the formative stages of development.
**Strengths**

We searched 9 of the most relevant bibliographic databases for medical and technology research for this review. No restrictions were placed on the year of publication, country of publication, journal, or study setting. Our study team consisted of multidisciplinary research and health care professionals with relevant expertise who provided direction at each review phase. This review was guided using an established framework proposed by Arksey and O’Malley [21].

**Limitations**

This review focused on simple voice- or text-based chatbots that engaged in 2-way communication with human users. This led to the exclusion of other forms of conversational agent technology (ie, embodied conversation agents, humanoid and social robots, wearable technology, IoT, virtual avatars, interactive voice assistants, etc) that may have resulted in the finding of additional development and engagement approaches that were not covered in our review. Our review excluded literature from conference proceedings, protocol papers, and other papers lacking an intervention. Moreover, although our proportionate agreement was 0.967 at the title and abstract screening stage, there was only “fair” agreement between reviewers (Cohen κ=0.309). This “fair” agreement between researchers highlights the challenges in reviewing a heterogeneous body of literature. With ongoing meetings and refinement of our inclusion and exclusion criteria, the Cohen κ statistic improved to an “almost perfect” agreement at the full-text review stage (Cohen κ=0.843). Additionally, due to the limited detail available within the included studies, our team could not conclusively assess patient involvement in chatbot development; greater attention to reporting patient involvement in chatbot development and testing in future research will help with this limitation. Finally, we acknowledge that scoping reviews have numerous shortcomings, including limitations of rigor and potential bias stemming from the absence of a quality assessment, among others [55]. However, the literature on chatbot technology remains highly heterogeneous at this time, and scoping review provided a systematic method to map the current state of the literature.

**Conclusion**

In conclusion, this review provides a menu of options that can be used for the nontechnical steps associated with chatbot development for interventions supporting lifestyle and wellness interventions. The identified study limitations hold promise to guide the inclusion of patient engagement and the improved documentation of the engagement and development of chatbots in future health care interventions. Given the importance of end user involvement in the development of digital technology, we hope that future research on chatbot development will take the opportunity to carry out a more systematic reporting of the chatbot development and implementation process and will actively engage patients as key members of the codevelopment process.
Acknowledgments

CS is a recipient of the 2022 Alberta SPOR Graduate Studentships in Patient-Oriented Research. Alberta SPOR Graduate Studentships in Patient-Oriented Research are jointly funded by Alberta Innovates and the Canadian Institute of Health Research. ND is granted funded by Mitacs and Cybera. AH is grant funded by Mitacs and Lupin Pharma Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy documentation.

References


Abbreviations

AI: artificial intelligence
DHI: digital health intervention
GRIPP2: Guidance for Reporting Involvement of Patients and Public
IoT: Internet of Things
PPI: patient and public involvement
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
WoZ: Wizard of Oz