Contents

Original Papers

An Luu, Truong Nguyen, Van Cao, Trinh Ha, Lien Chung, Trung Truong, Tung Nguyen Le Nhu, Khoa Dao, Hao Nguyen, Phan Khanh, Khanh Le, Luu Tran, Phung Nhat, Duc Tran, Yen Lam, Catherine Thwaites, Jacob Mcknight, Nguyen Vinh Chau, Jennifer Van Null, Vietnam ICU Translational Applications Laboratory (VITAL) 2

Usability of an App for Medical History Taking in General Practice From the Patients’ Perspective: Cross-Sectional Study (e47755)
Klara Albrink, Dominik Schröder, Carla Joos, Frank Müller, Eva Noack 12

A Novel Continuous Real-Time Vital Signs Viewer for Intensive Care Units: Design and Evaluation Study (e46030)
Shiming Yang, Samuel Galvagno, Neeraj Badjatia, Deborah Stein, William Teeter, Thomas Scalea, Stacy Shackelford, Raymond Fang, Catriona Miller, Peter Hu, VS viewer study group 27

User-Centered Design and Usability of a Culturally Adapted Virtual Survivorship Care App for Chinese Canadian Prostate Cancer Survivors: Qualitative Descriptive Study (e49353)
Karen Young, Ting Xiong, Rachel Lee, Ananya Banerjee, Myles Leslie, Wellam Ko, Quynh Pham 39

Effectiveness and User Perception of an In-Vehicle Voice Warning for Hypoglycemia: Development and Feasibility Trial (e42823)
Caterina Bérubé, Vera Lehmann, Martin Maritsch, Mathias Kraus, Stefan Feuerriegel, Felix Wortmann, Thomas Züger, Christoph Stettler, Elgar Fleisch, A Kocaballi, Tiborsz Kowatsch 50

An Phuoc Luu\(^1\), MPH; Truong Thanh Nguyen\(^2\), MD; Van Thi Cam Cao\(^2\), RN; Trinh Hoang Diem Ha\(^2\), RN; Lien Thi Thu Chung\(^2\), RN; Trung Ngoc Truong\(^2\), MD; Tung Nguyen Le Nhu\(^2\), MD; Khoa Bach Dao\(^2\), MD; Hao Van Nguyen\(^2\), MD; Phan Nguyen Quoc Khanh\(^1\), MD; Khanh Thuy Thuy Le\(^1\), MD; Luu Hoai Bao Tran\(^1\), MD; Phung Tran Huy Nhat\(^1\), BSc; Duc Minh Tran\(^1\), PMD; Yen Minh Lam\(^1\), MD; Catherine Louise Thwaites\(^1,3\), BSc, MBBS, MD; Jacob Mcknight\(^3\), BEng, MSc, DPhil; Nguyen Van Vinh Chau\(^1,2\), MD; Jennifer Ilo Van Nui\(^1,3\), BA, MA, PhD; Vietnam ICU Translational Applications Laboratory (VITAL)\(^4\)

\(^{1}\)Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
\(^{2}\)Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam
\(^{3}\)Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
\(^{4}\)See Acknowledgments

**Corresponding Author:**
Jennifer Ilo Van Nui, BA, MA, PhD
Oxford University Clinical Research Unit
764 Vo Van Kiet
Ho Chi Minh City, 70000
Vietnam
Phone: 84 362620124
Email: jvanhuil@oucru.org

**Abstract**

**Background:** Wearable devices have been used extensively both inside and outside of the hospital setting. During the COVID-19 pandemic, in some contexts, there was an increased need to remotely monitor pulse and saturated oxygen for patients due to the lack of staff and bedside monitors.

**Objective:** A prototype of a remote monitoring system using wearable pulse oximeter devices was implemented at the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam, from August to December 2021. The aim of this work was to support the ongoing implementation of the remote monitoring system.

**Methods:** We used an action learning approach with rapid pragmatic methods, including informal discussions and observations as well as a feedback survey form designed based on the technology acceptance model to assess the use and acceptability of the system. Based on these results, we facilitated a meeting using user-centered design principles to explore user needs and ideas about its development in more detail.

**Results:** In total, 21 users filled in the feedback form. The mean technology acceptance model scores ranged from 3.5 (for perceived ease of use) to 4.4 (for attitude) with behavioral intention (3.8) and perceived usefulness (4.2) scoring in between. Those working as nurses scored higher on perceived usefulness, attitude, and behavioral intention than did physicians. Based on informal discussions, we realized there was a mismatch between how we (ie, the research team) and the ward teams perceived the use and wider purpose of the technology.

**Conclusions:** Designing and implementing the devices to be more nurse-centric from their introduction could have helped to increase their efficiency and use during the complex pandemic period.

(JMIR Hum Factors 2024;11:e44619) doi:10.2196/44619

**KEYWORDS**

vital signs; wearable devices; action learning; technology acceptance model; TAM; COVID-19; user-centered design; wearables; remote monitoring; technology acceptance; oximeter
Introduction

The popularity of portable wearable technologies that monitor health has increased substantially over the past decade due to their perceived utility, relatively simple implementation, and immediate feedback [1]. Wearable technology is used in both personal and clinical settings, and more recently in the context of the COVID-19 pandemic for diagnosis, remote monitoring, and other applications in both inpatient and outpatient settings [1-3]. Using wearable devices for COVID-19 care can result in infection control by reducing the amount of time that health care workers (HCWs) are physically with patients and providing continuous monitoring of vital signs for the early identification and potential treatment of deteriorating patients [2]. Specifically, remote monitoring of oxygen saturation using wearable devices became increasingly common during COVID-19 in hospital settings [4,5].

Despite the potential benefits, there have been many challenges noted in implementing and using wearable devices during COVID-19, including technical, social, and political spheres [1]. Technical challenges often include battery life, Wi-Fi or Bluetooth connections, and device communication. A few examples of social challenges are patients lacking technological confidence (e.g., in older patients) and repeated device alerts or continuous monitoring making patients nervous, while political challenges could include regulatory issues for expanding the approval of devices for COVID-19–related medical situations [1]. Many of these challenges may be enhanced in low- and middle-income countries (LMICs), while the need for such integration is crucial, especially during pandemic situations [6,7].

There are several studies exploring the technical challenges of integrating wearable devices in trial settings during COVID-19 (e.g., see [8]), but there is a lack of research surrounding the acceptability of such devices within these contexts and how attitudes may impact actual use [9]. Portable wearable devices could be a potential solution to allow for continuous monitoring of vital signs remotely and affordable for COVID-19 wards in LMIC settings; however, while advantageous, these devices cannot meet their full potential if the users do not agree to use them or realize their potential value [10]. Understanding user perceptions and needs as well as the context in which the technology is implemented is crucial for successful implementation [1]. User-centered approaches stress the importance of integrating both human factors and technical factors [11] while also paying attention to avoid excluding certain populations in the design [7]. User-centered approaches have been cited as a “critical success factor” in a variety of health-related technology projects [12].

From August to December 2021, when COVID-19 cases were increasing more rapidly than at any time previously in Ho Chi Minh City, Vietnam, there was an opportunity to integrate a prototype wearable device and monitoring system into the COVID-19–designated wards at the Hospital for Tropical Diseases (HTD). At this time, the HTD was overwhelmed with patients with COVID-19 and we needed to deploy something urgently that could help. Using pragmatic methods during the rollout of the device, we describe stakeholders’ use of the wearable device, aspects of acceptability, and under which circumstances its use would be most beneficial for improving the care of patients with COVID-19. The primary objective of this work was to support the implementation process of the wearable device in the hospital to improve patient care during a catastrophic period of the COVID-19 pandemic in Ho Chi Minh City, Vietnam.

Methods

Study Setting

This work took place within a larger project called the Vietnam ICU Translation Application Laboratory (VITAL) at the Oxford University Clinical Research Unit (OUCRU) and HTD. The goal of VITAL is to design and implement innovative technologies to improve patient care within the intensive care unit (ICU) at the HTD, with a longer-term goal of expanding these technologies regionally. In addition to the clinical and technological studies, there is an ethnographic study to explore the sociotechnical contexts of the ICU at the HTD and within ICUs in Vietnam more broadly. The VITAL multidisciplinary team was in place at the start of the COVID-19 pandemic.

In the first 100 days of the COVID-19 pandemic, Vietnam rapidly implemented a variety of public health measures resulting in relatively few cases and zero deaths [13]. Since that time, there were a few concentrated outbreaks (for example, in Da Nang in July 2020 and December 2020 in northern Vietnam). In May 2021, the cases started to increase on a countrywide basis, and by August 2021, the hospitals began to fill with patients with COVID-19. It was within this pandemic context that the wearable device was implemented at the HTD, and the VITAL study teams worked together throughout to improve its implementation.

The wearable device was selected by the company and was already integrated into a locally developed platform based on an available application programming interface, licensing, and availability. The device was medical grade and measured heart rate and blood oxygen levels, similar to a pulse oximeter. The wearable device was battery powered and each one connected to a tablet that was kept at the patients’ bedside. The tablets had a 3G or 4G sim card and sent the data to a cloud where multiple patients’ data were viewable by HCWs outside the patients’ rooms and isolation area.

Study Design

The aim of this work was to support the ongoing implementation of the wearable device rather than to follow a predefined, replicable study protocol, as would be used in trial settings, for example. Therefore, the work here describes the pragmatic rollout of the device. We used an action learning approach, including integration of multiple methods to assess the use and acceptability of the wearable device [14]. Action learning approaches rely on an iterative process of assessing local contexts, learning from relevant stakeholders, and using the information to improve an implementation or further develop a technology specific to the context [15,16]. As the wearable device started to be implemented in the HTD wards, our team
of HCWs, social scientists, and technology developers took the opportunity to work together to inform the implementation. Therefore, we adapted the methods as the situation changed and more insights were gained [14].

**Participants**

Potential participants included the HCWs from the HTD who were using the device in the wards during the implementation and corresponding ward heads. We estimated that a total of 30 doctors and 60 nurses would have worked in the wards where the wearable device was implemented and potentially used it in some form; therefore, we planned to recruit participants from this larger sample.

**Data Collection Methods**

**Informal Discussions and Observations**

We used an iterative process of engaging in informal discussions coupled with sense-checking discussions and observations during the implementation period. The informal and sense-checking discussions and observations were conducted with the team who was working directly in the wards, as well as with head nurses from the wards where the wearable device was being implemented. The informal discussions and observations were conducted during the implementation of the device.

**Feedback Survey Form**

We created the feedback form based on the components of the technology acceptance model (TAM) to assess the use and acceptance of the device. The TAM is used in a variety of disciplines to determine how individuals accept (or not) and use (or not) a given technology. Davis [17] developed this model based on components from the theory of reasoned action [18] and it consists of the following variables: use motivation (with perceived ease of use and perceived usefulness) and behavioral intention [17,19]. The model suggests that an individual will accept the use of a technology (ie, their behavioral intention) based on their perception of the technology’s usefulness and ease of use. Perceived usefulness refers to the perception that using the technology will enhance one’s work; for example, the wearable device will provide physicians and nurses some advantages (eg, remote monitoring). Perceived ease of use refers to the perception that the use does not add more work or effort to the work that could be enhanced; for example, using the wearable device will not increase nurses’ workload, despite its utility and simplicity [17]. The TAM framework was expanded twice to include attitudes as well as several other external factors [20]. The use of the TAM in health research has shown how perceived usefulness and perceived ease of use relates positively to attitude and behavioral intention [21]. The TAM has been criticized for being insensitive to the context or social factors, being simplistic, and following an assumption that users are rational decision makers, when indeed other factors play into decision making [22-24]. We used the TAM framework for its simplicity and because the categories of perceived usefulness and perceived ease of use were of relevance, but we also integrated other data collection methods alongside it to counter these limitations to some extent.

Based on the components of the TAM, we included 23 questions related to usefulness (n=5), ease of use (n=5), attitude (n=5), and behavioral intent (n=8) [25]. We asked these questions using a 5-point Likert scale (with scores of 5 being more favorable). We also added 2 open-ended questions and collected a variety of relevant demographic information (Multimedia Appendix 1). We piloted the tool in both English and Vietnamese and adjusted the form as needed. We used Google forms for electronic self-completion of the form and offered paper forms for hand-written self-completion. We explained the feedback form to the ward staff during team meetings and provided the link. The feedback form was distributed and completed in Vietnamese. We kept the feedback form link open for 7 weeks in total and started data collection after the implementation had been integrated into the wards so that users would have had experience using the device.

**User-Centered Workshop**

We held a user-centered workshop with a selection of HTD ward staff to explore user needs and ideas for development in more detail. Because we already had the technology and knew the spaces where implementation would be held, we followed an adapted version of the process described by Cooper et al [26]. With this approach, the workshop participants and facilitators set the scene as a busy COVID-19 ward during the peak of the pandemic. Then, the facilitators described the shells of users (personas), including a nurse and a doctor persona shell, and we had the workshop attendees describe who they imagined the nurse and doctor to be, as well as their behaviors and needs and the values each user group would find most essential. We based the conversation on the wearable technology that the participants had already used. Then, the group discussed solutions to the issues identified [26].

**Data Analysis**

Using the principles of action learning, we integrated the responses from informal discussions and observations into subsequent data collection, as well as summarized the content and grouped it into themes. For the analysis of the feedback survey form, we calculated mean scores for each variable and compared scores by profession. For the open-ended survey questions, we used content coding to summarize the responses topically. We presented the demographic data descriptively. We documented the responses from the user-centered design workshop as notes and summarized the results into main themes.

**Ethical Considerations**

In this paper, we are describing the processes that occurred as part of the development and implementation of a monitoring system; therefore, the work did not require ethics approval. Prior to the initiation of any activities, we held a meeting with ward heads to describe the work in more detail and obtain their agreement.

**Results**

**Device Implementation Within the HTD Context**

The wearable device was implemented in 3 wards starting in August 2021, including the adult ICU, Ward A, and Ward E.
We describe the implementation over a 5-month period from August to December 2021. During this period, these wards changed from COVID-19–designated and then back again to routine patient care settings, depending on the number of patients. Although the HTD was one of the COVID-19–designated hospitals, throughout the pandemic they offered routine patient care for specific diseases (eg, tetanus).

In addition to the rapidly changing physical spaces, the hospital management quickly deployed remote monitoring capacity using existing closed-circuit television cameras as a temporary solution to monitor very sick patients from outside the patients’ rooms. The remote monitoring was useful as it allowed for multitasking and prevented nurses and doctors from checking on patients more routinely in person. The hospital wards were at capacity during the study period. Prior to the pandemic, however, it was not unusual for the wards at the HTD to often be at maximum patient capacity. For example, in the adult ICU or during the rainy season, the number of dengue patients increases dramatically and the wards tend to be full.

Also, the workflow was organized differently during the pandemic period. Instead of nurses taking care of a few specific patients for the whole shift, 2 nurses and 1 doctor would instead go into the ward (in full personal protective equipment) as a team for 3 hours at a time while the other 2 nurses on shift completed admin work in the office. This meant that more coordination was needed, and often the team with the patients “needed someone else to be [their] memory” as it was not easy to remember everything about all patients. The health care team’s workload, especially that of the nurses, ended up being more extensive for many reasons. One important reason is that, because of COVID-19 restrictions, there were also no families allowed in the wards who would help to look after patients in non–COVID-19 times; therefore, the majority of the care was left to the nurses. The patients were also more severely ill than previously in these wards and required more care by fewer staff.

**Device Use and Acceptability**

When we first distributed the feedback form, out of 90 potential participants, only 22 completed the survey (19 electronic and 3 paper forms), and 1 person stated that they did not use the technology and therefore no responses were recorded for that participant. Of the 21 respondents who completed the feedback form, 48% (n=10) were doctors and 48% (n=10) were nurses, with 52% (n=11) of the participants coming from Ward E (Table 1).

Overall, when assessing the TAM variables, the mean (SD) scores ranged from 3.6 (0.8) for perceived ease of use to 4.4 (0.6) for attitude, with behavioral intention (mean 3.9, SD 0.6) and perceived usefulness (mean 4.2, SD 0.7) scoring in between. Those working as nurses scored higher on perceived usefulness, perceived ease of use, attitude, and behavioral intention than did physicians (Table 2).

When asked, as an open-ended question, why participants would or would not use the wearable device in the future, of the 19 responses inputted, 15 participants wrote that they would use the system because of its convenience and usefulness in monitoring patients. However, in 2 of those responses, they also added comments that the device had limited perceived accuracy and transmission problems. Of the remaining 4 participants, 1 participant simply stated that the monitor was still in use, 2 participants wrote that they did not use the system anymore due to job location changes, and 1 participant wrote a few sentences about why the wearable device is not the “best choice,” highlighting its limited battery life, how the system had become additional work for the already overworked staff, and how it is not yet completely implemented.

**Table 1.** Demographic characteristics of the survey respondents (n=21).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>13 (62)</td>
</tr>
<tr>
<td>Men</td>
<td>8 (38)</td>
</tr>
<tr>
<td><strong>Age (years), median (IQR)</strong></td>
<td>35 (30-38)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Doctor</td>
<td>10 (48)</td>
</tr>
<tr>
<td>Nurse</td>
<td>10 (48)</td>
</tr>
<tr>
<td>Other: nurses’ aid</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Primary ward during the implementation phase, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Adult intensive care unit</td>
<td>6 (29)</td>
</tr>
<tr>
<td>Ward A</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Ward D</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Ward E</td>
<td>11 (52)</td>
</tr>
</tbody>
</table>
Integrating User Perceptions for Improved Implementation

As part of the action learning process, we supplemented the feedback form results with data from the observations and informal discussions during the 5-month period. There were 3 main observations. First, there was a mismatch between how we (ie, the research team) and the ward teams perceived the use of the technology. We quickly realized, from our observations and from informal discussions with the implementation team, that many of the nurses either did not use the wearable device or did not think that they used it even if they used it in some aspect (eg, connecting the device for the patients or changing batteries). Even after we clarified what we meant by “use,” there were still not additional participants who filled in the feedback form because they felt like they did not use the technology.

Second, the ward teams had varying perceptions of the technologies that are routinely implemented by the OUCRU team in the HTD wards as part of research projects. We heard from informal discussions with colleagues that the nurses assumed the wearable devices were from a research project, as is often the case with OUCRU projects, and therefore the nurses, in particular, ignored the device even if they had some role in its use. They did not see its potential benefit.

Finally, in order to make the device more useful for the ward staff, we realized during the meetings and informal discussions with the team that we needed to make the implementation and use of the device more “pro-nurse,” meaning we would need to emphasize how the device and its data were also useful and relevant to them. When discussing with the head nurse, the data were only displayed on the main screen in the staff room for one department. One suggestion was to move the tablet to the wall so that the nurses and others in the room (including the patients) could potentially see their vital signs. Because the devices and corresponding data were not in sight, it was easy to think that it was not relevant for the nurses and made it easier for them to ignore the device while with the patients.

User-Centered Design Workshop

With the information we had learned from the informal discussions, observations, and feedback form, we held a follow-up workshop on January 17, 2022, to discuss how we could make better use of the technology in the wards in COVID-19 situations in the future. The attendees included 2 doctors (1 man and 1 woman) and 3 nurses (2 women and 1 man). The participants discussed the behaviors and needs of the nurse and doctor persona. For both roles, the needs centered on having equipment and improved coordination. The nurses also mentioned more training needs, while the doctors’ needs were about the accuracy of monitoring (Textboxes 1 and 2).

There were 3 main value prop themes, including medical, technical, and patient themes. For medical aspects, the attendees discussed how the device should be able to provide highly accurate data, with appropriate alarms and cut-offs. For the technical theme, the device and software should be simple to connect and use, with a long battery life and stable connections during charging or switching devices. The display should be large and clear, and the data should be stored for a long period of time (ie, 7-10 days). Finally, for the patient theme, the device should be comfortable for the patients to wear to avoid them removing it.

There were several solutions discussed in the group to improve the use and efficacy of the wearable device (Table 3). Solutions included improving the credibility of the data, ideas to improve the ease of use, ways to make the alarms more consistent, and ideas for more ideal placement. One very specific issue that the group mentioned was that the alarms went off too much on the large display and the alarms were always red or black and blinking, and it was difficult to know if the device was turned off (due to patient discharge) or actually disconnected, which would require an intervention. The solution was to refresh the devices; however, if the alarms were excessive and not always indicating a real issue, trust in the device would remain low, so this was an important priority. They also suggested that the alarms and display on the tablet should be the same as the big screen, as they preferred screen consistency.

Another in-depth discussion was about moving the tablets to the walls and having the device plugged in all the time, which would solve the battery issues. They felt that the tablet could be set up on the wall but that brought up other issues about how to keep the device and watch safe after use. For some of the topics, the group used features of another wearable device that they had used in the wards in the past to inform their solutions (eg, device graphs and a line on the device for finger placement).

Table 2. Mean technology acceptance model (TAM) scores by variable. The maximum score was 5.

<table>
<thead>
<tr>
<th>TAM variable</th>
<th>All participants, mean (SD)</th>
<th>Nurses, mean (SD)</th>
<th>Doctors, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness</td>
<td>4.2 (0.7)</td>
<td>4.3 (0.8)</td>
<td>4.1 (0.6)</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>3.6 (0.8)</td>
<td>3.8 (0.7)</td>
<td>3.4 (0.8)</td>
</tr>
<tr>
<td>Attitude</td>
<td>4.4 (0.6)</td>
<td>4.6 (0.6)</td>
<td>4.2 (0.6)</td>
</tr>
<tr>
<td>Behavioral intention</td>
<td>3.9 (0.6)</td>
<td>4.0 (0.6)</td>
<td>3.7 (0.7)</td>
</tr>
</tbody>
</table>
Textbox 1. Behaviors and needs of the nurses.

Participant: Nurse Van is a 36-year-old woman. She is an administrative nurse and has a management job. She likes to have fun and has a family and 2 children. She is also responsible for bringing the kids to school and back.

Behaviors:
- Visit and provide direct patient care and monitor vital signs
- Carry out medical orders (ie, medications, blood tests, and nutrition)
- Assess, monitor, and hand over patients
- Work night duty
- Night shifts inform doctors on vital signs as prescribed

Needs:
- Equipment (eg, to measure blood pressure, temperature, oxygen levels, and heart rate)
- Training on diseases
- Teamwork and coordination

Textbox 2. Behaviors and needs of the doctors.

Participant: Doctor Huong is a 30-year-old woman. She is flexible and very active. She is not married and has no children and currently lives in a hotel. She is on night shift every 4 nights, and at times she visits her home in another town in Ho Chi Minh City, which is far from the Hospital for Tropical Diseases.

Behaviors:
- Prescribe medications
- Update medical records
- Perform examinations and change treatments
- Data entry
- Check vital signs in patient rooms (with a portable monitor that they move around) for examination and to detect abnormalities

Needs:
- Equipment (eg, monitors)
- Coordination with nurses (progress: medical records)
- Re-evaluation and working with other doctors
- Accuracy of vital sign monitoring

Table 3. Solutions for improvement.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Specific solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data credibility</td>
<td>Adding a graph for signal strength</td>
</tr>
<tr>
<td>Ease of use</td>
<td>• Adding a finger placement mark on the device</td>
</tr>
<tr>
<td></td>
<td>• Increasing the font size on the watch and tablet</td>
</tr>
<tr>
<td></td>
<td>• Tablets should be fixed on the wall</td>
</tr>
<tr>
<td>Alarms</td>
<td>• Reduce the alarm colors and blinking on the screen</td>
</tr>
<tr>
<td></td>
<td>• Use the same display on the screen and the tablets for consistency</td>
</tr>
<tr>
<td></td>
<td>• Refresh the tablets for more accurate alarms</td>
</tr>
<tr>
<td>Battery issues</td>
<td>Keep the tablet plugged in</td>
</tr>
<tr>
<td>Device placement</td>
<td>Placement on the wall (but only with an increase in font size)</td>
</tr>
</tbody>
</table>
Discussion

The HTD and OUCRU teams, along with the technology company, rolled out the wearable device in an extremely complex pandemic situation with a prototype system. In the end, the team used the device on over 100 patients. We assessed the usability and acceptance of the device over the implementation period when COVID-19 cases were peaking in the hospital and into the period when the COVID-19 cases were reducing. Similar to the literature on the topics, we found that the importance of understanding the users and their experiences using the device was crucial to get the most use out of these technologies.

There was a mismatch between our perception of who was using the device and those who thought that they were using or benefiting from the device on the ground. From the start, the device was designed and set up with doctors in mind, but in practice, the nurses’ roles and use were overlooked, even though they could also routinely use and benefit from the device. In our study, we found that the nurses who filled in the feedback form, on average, had slightly higher scores on 3 of the 4 TAM domains (ie, perceived usefulness, attitude, and behavioral intention), while the doctors, on average, scored the perceived ease of use slightly higher than the nurses. We know from the challenges with acquiring feedback that many nurses did not feel that they used the device even though they had some role in the device set-up and monitoring. Designing the device to be more nurse-centric from the early phases could have helped to increase the efficiency and definition of who is meant to use it. In the future, it is important to consider that the way the device is used might be dependent on the form of its use (eg, for triage, use in a pandemic emergency, or routine hospital use). We recommend the involvement of staff who could benefit from the technology, especially nurses in the hospital context, in the full implementation process. This could help to avoid mismatches in the perceptions of who the users are and who could and should benefit from the new technology. Research on integrating wearable devices during COVID-19 in Singapore also highlighted that device simplicity would encourage its use and the importance of making the technology fit into the current environment while not increasing or disrupting workflows [27].

The trust in the device and its data was an issue brought up several times during the implementation and feedback sessions. There are a variety of potential explanations for inconsistent data (eg, incorrect device placement or averaging of data); however, it reduced the credibility of the device for both doctors and, importantly, nurses. Data concerns about technology in clinic settings has been noted in other studies. For example, Faria et al [28] found that study clinicians reported that 36% of the data from a remote monitoring project were “invalid” for a variety of reasons, including low literacy of the patients and complexity of the device. Involvement from users from the beginning of the design and implementation process is crucial for design purposes but also to build trust and confidence in the devices [11]. While this project took place during COVID-19, which is a very specific circumstance, the broader findings resonate with research conducted prior to COVID-19 that focused on the implementation and scaling up of digital health technologies in LMICs. The recommendations also included integration of end-user feedback and engagement with all stakeholders throughout the design and implementation process [12].

There are limitations to this work. First, we did not collect data on the clinical worth or the accuracy of the data transmitted from the devices. Second, we focused on feedback from only health care staff (ie, doctors and nurses), and from only a subset of those who perceived that they used the device, which may have excluded some users and limited the overall sample size. We did not include patients who could also inform device acceptance, especially if used in noncritical cases where patients are moving around and conscious. Finally, the implementation setting for this work is not typical of other hospital settings in Vietnam or possibly other LMICs, as the HTD is a large referral hospital with an international research institute attached to it.

In anticipation of future (novel) pandemic situations or integration of wearable technologies into a range of clinical settings more broadly, it is important to fully understand if and how the wearable devices could be used more effectively by doctors, and importantly, nurses in the wards, for monitoring of deteriorating patients, especially in LMICs where resources are already stretched. Using an action learning approach during the implementation process highlights the importance of integrating user perspectives, ideas, and solutions into development and design.

Acknowledgments

We would like to acknowledge Nguy n D i Ph and Nguy n Th Tin from the Hospital for Tropical Diseases and Nguy n Th Di m Trinh and D ng Ph ng Th o from the Oxford University Clinical Research Unit for their support and commitment in operating the remote monitoring system.

The members of the Vietnam ICU Translational Applications Laboratory (VITAL) group are as follows: Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam: An Phuoc Luu, Chanh Quang Ho, Duc Hong Du, Duc Minh Tran, Dung Thi Phuong Nguyen, Giang Thi Nguyen, Hai Bich Ho, Hien Van Ho, Hung Manh Trinh, Huy Quang Nguyen, Khanh Nguyen Quoc Phan, Khoa Dinh Van Le, Kien Trung Dang, Lam Khanh Phung, Lieu Thi Pham, Ngoc Thanh Nguyen, Nhat Tran Huy Phung, Phuong Thanh Le, Quyen Than Ha Nguyen, Thanh Thi Le Nguyen, Thy Bui Xuan Doan, Trieu Trung Huynh, Trinh Huu Khanh Dong, Van Minh Tu Hoang, Van Thi Thanh Ninh, Vuong Lam Nguyen, Yen Minh Lam, Sayem Ahmed, Joseph Donovan, Ronald Geskus, Evelyne Kestelyn, Angela Mcbride, Guy Thwaites, Louise Thwaites, Hugo Turner, Jennifer Ilo Van Nuil, and Sophie Yacob; Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam: Tam Thi Cao, Thuy Bich Duong, Duong Thi Hai

https://humanfactors.jmir.org/2024/1/e44619

JMI Hum Factors 2024 | vol. 11 | e44619 | p.8
(page number not for citation purposes)
Ha, Nghia Dang Trung Ha, Chau Buu Le, Thu Ngoc Minh Le, Thao Thi Mai Le, Tai Thi Hue Luong, Phu Hoan Nguyen, Viet Quoc Nguyen, Nguyen Thanh Nguyen, Phong Thanh Nguyen, Anh Thi Kim Nguyen, Hao Van Nguyen, Duc Van Thanh Nguyen, Chau Van Vinh Nguyen, Oanh Kieu Nguyet Pham, Van Thi Hong Phan, Qui Tu Phan, Tho Vinh Phan, and Thao Thi Phuong Truong; University of Oxford, Oxford, United Kingdom: David Clifton, Mike English, Shadi Ghiassi, Heloise Greeff, Jannis Hagenah, Ping Lu, Jacob McKnight, Chris Paton, and Tingting Zhu; Imperial College London, London, London, United Kingdom: Pantelis Georgiou, Bernard Hernandez Perez, Kerri Hill-Cawthorne, Alison Holmes, Stefan Karolcik, Damien Ming, Nicolas Moser, and Jesus Rodriguez Manzano; King’s College London, London, United Kingdom: Alberto Gomez, Hamideh Kerdegari, Marc Modat, and Reza Razavi; ETH Zurich, Zurich, Switzerland: Abhilash Guru Dutt, Walter Karlen, Michaela Verling, and Elias Wicki; The University of Melbourne, Melbourne, Australia: Linda Denney and Thomas Rollinson.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Study tools in English and Vietnamese.

[PDF File (Adobe PDF File), 240 KB - humanfactors_v11i1e44619_app1.pdf]

**References**


6. Local production and technology transfer to increase access to medical devices addressing the barriers and challenges in low-and middle-income countries. World Health Organization. URL: https://www.who.int/publications/i/item/9789241504546 [accessed 2023-12-07]


Abbreviations
HCW: health care worker
HTD: Hospital for Tropical Diseases
ICU: intensive care unit
LMIC: low- and middle-income country
OUCRU: Oxford University Clinical Research Unit
TAM: technology acceptance model
VITAL: Vietnam ICU Translation Application Laboratory

Edited by A Kushniruk; submitted 28.11.22; peer-reviewed by K Liu, M Harris, J Claggett; comments to author 15.04.23; revised version received 15.06.23; accepted 20.11.23; published 05.01.24.

Please cite as:
URL: https://humanfactors.jmir.org/2024/1/e44619
PMID:38180799

©An Phuoc Luu, Truong Thanh Nguyen, Van Thi Cam Cao, Trinh Hoang Diem Ha, Lien Thi Thu Chung, Trung Ngoc Truong, Tung Nguyen Le Nhu, Khoa Bach Dao, Hao Van Nguyen, Phan Nguyen Quoc Khanh, Khanh Thuy Thuy Le, Luu Hoai Bao Tran, Phung Tran Huy Nhat, Duc Minh Tran, Yen Minh Lam, Catherine Louise Thwaites, Jacob Mcknight, Nguyen Van Vinh Chau, Jennifer Ilo Van Nuil, Vietnam ICU Translational Applications Laboratory (VITAL). Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 05.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Usability of an App for Medical History Taking in General Practice From the Patients’ Perspective: Cross-Sectional Study

Klara Albrink; Dominik Schröder, MSc; Carla Joos; Frank Müller, MSc, Dr med; Eva Maria Noack, PhD
Department of General Practice, University Medical Center Göttingen, Göttingen, Germany

Corresponding Author:
Eva Maria Noack, PhD
Department of General Practice
University Medical Center Göttingen
Humboldtallee 38
Göttingen, 37073
Germany
Phone: 49 551 39 65745
Email: evamaria.noack@med.uni-goettingen.de

Abstract

Background: A future shortage of physicians, especially in general practice, will result in an increasing workload for health care providers as a whole. Therefore, it is important to optimize patient-encounter processes to increase time efficiency related to visits. Utilizing digital tools to record patients’ medical histories prior to a consultation offers great potential to achieve this goal. The collected information can be stored into the practice’s electronic medical record, allowing for the general practitioner to review structured information of the patients’ complaints and related medical history beforehand, thereby saving time during the encounter. However, the low usability of new digital developments in this setting often hinders implementation.

Objective: The aim of this study was to evaluate the usability of an app designed for medical history taking in general practice to capture the patients’ perspective.

Methods: Between November 2021 and January 2022, we recruited 406 patients with acute complaints in one out-of-hour urgent care and seven general practice clinics. These study participants used the app during their waiting time and subsequently assessed its usability by completing the System Usability Scale (SUS), a robust and well-established 10-question survey measuring the perceived usability of products and technologies. Additionally, we collected general participant information, including age, sex, media usage, health literacy, and native language. Descriptive and inferential statistics were applied to identify patient characteristics associated with low or high SUS scores.

Results: We analyzed data from 397 patients (56.7% female, 43.3% male). The mean total SUS score was 77.8 points; 54.4% (216/397) of participants had SUS scores of 80 points or higher, indicating high usability of the app. In a multiple linear regression predicting SUS score, male sex and higher age (65 years or older) were significantly negatively associated with the SUS score. Conversely, a higher health literacy score and German as the native language were significantly positively associated with the SUS score.

Conclusions: Usability testing based on the SUS anticipates successful implementation of the app. However, not all patients will easily adapt to utilizing the app, as exemplified by the participants of older age in this study who reported lower perceived usability. Further research should examine these groups of people, identify the exact problems in operating such an app, and provide targeted solutions.

Trial Registration: German Clinical Trials Register World Health Organization Trial Registration Data Set DRKS00026659; https://trialsearch.who.int/Trial2.aspx?TrialID=DRKS00026659

(JMIR Hum Factors 2024;11:e47755) doi:10.2196/47755

KEYWORDS
digitization; application software; usability; mHealth; history of present illness; medical history taking
**Introduction**

As in many countries, demographic change is becoming evident in the German health care system, resulting in more complex, multimorbid patients [1] and a shortage of physicians [2]. Moreover, the proportion of older people in the population is rising steadily [3] and people tend to use medical services at a higher rate as they increase in age [4]. In Germany, one group that is particularly affected by this development are general practitioners (GPs) who are the first point of contact for patients requiring medical care and serve as the “gatekeepers” in the German health care system [5]. Approximately 80% of the German population aged 18 years and older are treated by a GP at least once a year [6]. A considerable number of GPs will retire in the upcoming years, resulting in 11,000 GP vacancies expected by 2035. These vacancies will disproportionately impact structurally weak and rural areas [7,8]. Without a sufficient workforce to replace the retired GPs and meet the greater demand for physicians, remaining GP workloads are expected to increase significantly within the next decade [9]. These developments challenge the German health care system at various levels and require attention to address the following key issues: future financing, improving allocation of resources, ensuring access to care, increasing efficiency and effectiveness of health care provision, and strengthened collaboration between providers [10].

To streamline patient care in the upcoming years, it is of importance to optimize patient-encounter processes to increase time efficiency related to visits. In this respect, digital tools offer great potential to support GPs in patient management, documentation workload, and the collection of medical history before consultation.

Digital tools designed to collect patients’ medical history can ensure that information is always collected and documented thoroughly in a structured manner and with consistent quality. As many conditions can be diagnosed via a thorough medical history [11,12], these tools can be helpful in maintaining quality of care when time constraints may lead to an otherwise superficial medical history.

As part of our project titled “Digitally assisted information acquisition before medical consultation” (DASI), we developed an app for medical history taking in general practice settings. The app is used by the patient prior to the medical consultation, which could be either in the waiting area or at home. The collected information can be stored in the practice’s electronic medical record and eventually be transferred to the individual electronic patient file, which statutory health insurers in Germany have been entitled to use since 2021 [13]. In the electronic patient file, patient data such as medical reports, X-rays, immunization records, and other medical data can be stored and shared among medical providers involved in the care of a particular patient by using the telematics infrastructure [14].

One advantage of the app is that the GP can review structured information of the patients’ complaints and related medical history before the encounter. This is particularly helpful for patients that are unknown to the provider, those with many complaints, or those who have a comprehensive medical history. These situations are especially prevalent in out-of-hour urgent care practices. Furthermore, the tool might help patients to reflect on their conditions and enable them to better address their needs when seeing the provider. In this way, we expect that the limited consultation time can be used more efficiently.

Despite Germany’s progress in digitalization within the health sector, concerns remain about the limited usability of new digital tools, hindering their full implementation. More than half of German practices see low usability as a strong obstacle to digitalization [15]. The evaluation of a digital tool’s perceived usability is of special interest as it is a key determinant of performance for end users. Therefore, the aim of this study was to assess the usability of the app from the patients’ perspective and to identify features in need of improvement. The broader aim is to ensure that the app is suitable for implementation in everyday practice, considering that GPs treat a broad range of patients of all ages and various educational and cultural backgrounds [16].

**Methods**

**Study Design and Recruitment**

This was a cross-sectional study conducted in Germany in one out-of-hour urgent care practice and seven GP practices to assess the usability of an app designed for medical history taking in general practice settings.

**Software and Hardware**

The app was developed to take a medical history based on general medical complaints directly from the patients. While there are no international standards for the composition of a standardized patient history, this app was developed based on guidelines and health literature by medical experts from aidminutes GmbH (Hamburg/Buchholz in der Nordheide, Germany). For this study, the content and query structure were further refined for primary care (general practice and out-of-hour practices) by aidminutes GmbH in collaboration with experienced researchers from the Department of General Practice at the University Medical Center Göttingen, Germany. The app was designed to be used by patients in the waiting room before they see the doctor. Patients select one or several complaints and are then guided through a symptom-related questionnaire. In the sense of a branching logic, the app is adaptive to patient responses, which trigger further specific questions about the selected key complaints (eg, how and when a symptom started). Patients are also asked about preexisting conditions, previous treatments and surgeries, current medication, living habits, and chronic conditions in the family history. Information such as biological sex, height, weight, age, as well as the subjectively perceived severity of the complaints are inquired from all patients. More details can be found in the published study protocol [17].

The app was designed to be intuitive for the user such that no prior knowledge or any kind of instruction for its use is necessary. The user interface was designed to be simple to follow and only one question is asked per screen. As the app is operated in the waiting area, sound and video output of an earlier version [18] was omitted due to data protection. The questions
are phrased in plain language; medical terminology is avoided or otherwise explained. The questions are substantially comprised by single-choice or multiple-choice questions that can be answered by tapping but also include several data fields (for age, height, and weight) and slider-type questions (Figure 1). The color scheme was designed to ensure reading accessibility for patients who may be color blind. A zoom function can be used for users who may experience visual impairment.

**Figure 1.** Screenshots of the app for medical history taking in general practice showing different types of questions: (A) single-choice question; (B) multiple-choice question; (C) hybrid question (ie, patients can either select several options or negate all of them); (D) slider for questions including a ranking between items (depicted here as “How sick do you feel?”); (E, F) data entry field (here: “Please enter your age”); and (G, H) selection of a body region on a figure (depicted in figure: “Please mark on the figure where you are suffering from the problems”).

As this is a web-based app, it relies on a permanent internet connection. For this study, the app ran on an iPad Mini 4 (Apple Inc, Cupertino, CA, USA) held in an upright position. Tablets were equipped with haptically and visually inconspicuous cases (dark grey polyurethane leather outside and microfiber inside).

**Setting**

In Germany, GP practices aim at providing preventive, acute, and rehabilitative health care with long-lasting patient-doctor relationships. Out-of-hours urgent care practices provide urgent medical care for acute but not life-threatening cases when other practices are closed. Urgent care practices are often staffed with doctors of various specialties and an established relationship of care between the patients and doctors is not common. These aspects can lead to challenges in efficiently obtaining an accurate medical history and identifying serious health problems. Although the app was designed for general practice, it is also suitable to be used in out-of-hour urgent care practices.

**Data Collection**

The recruitment of patients was carried out by three study nurses and took place from November 22, 2021, to January 12, 2022. Patients were approached by study nurses in the waiting room of the respective practices before seeing their GP.

Patients meeting the following criteria were eligible to participate in the study: (1) seeking care in a participating practice because of acute somatic and/or psychological complaints, (2) at least 18 years old, and (3) consenting to participate in the study. Patients meeting the following exclusion criteria could not participate in the study: (1) younger than 18 years old (legal minor), (2) patients in an apparent emergency, (3) patients who required immediate medical treatment, and (4) patients who were unable to provide consent.

After the study nurses obtained written informed consent, a tablet on which the app was run on was handed over to the study participants. Participants used the app to report their medical history without an introduction on how to navigate the app. Once finished, they were asked to answer questions on personally perceived usability, media usage, and further
sociodemographic data, which were digitally attached to the medical history—taking document. The study nurse in charge was present to observe any problems study participants may have had with using the app and was available to answer questions about the app’s content and usability if specifically requested. Data were collected in an anonymized format without any personal information (eg, name or address) linking the results to each study participant. More detailed information on the data collection can be found in the study protocol [17].

Ethical Considerations
The Medical Ethics Committee of the University Medical Center Göttingen approved the study (approval number 26/3/21). A written informed consent form was collected from all patients before their inclusion in the study. Participating in the study was voluntary for patients. Patients could withdraw from participation without giving a reason at any time before they had completed the survey. Subsequently, their data could no longer be deleted because it could not be traced back to the individual.

Measures
The main outcome “usability” was measured using the System Usability Scale (SUS) [19], a commonly used instrument for this purpose [20]. The SUS was developed based on Standard ISO 9241-11 [21], in which usability is measured by the three main attributes of “effectiveness,” “efficiency,” and “satisfaction” [22,23]. Compared to other instruments, the SUS offers several advantages: (1) it can be analyzed quickly, (2) it is relatively easy to understand by academics from other disciplines [24], (3) it contains only 10 statements for easy completion, and (4) it can be used to evaluate almost any type of user interface [25]. We used the translated and validated German version of the SUS [26] and modified the statements to suit our purpose (see Multimedia Appendix 1).

The SUS consists of 10 statements (Table 1), where statements 1, 3, 5, 7, and 9 are positively connoted and statements 2, 4, 6, 8, and 10 are negatively connoted [19]. The scores for these statements are therefore inverted when calculating the sum. The raters decide on the extent to which they agree or disagree to these statements on a 5-point Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree). The final sum score is multiplied by 2.5, resulting in a score range of 0-100 with higher scores indicating better usability [19].

Lewis and Sauro [27] developed a curved grading scale for SUS scores by comparing more than 200 industrial usability studies and using the percentile ranges, resulting in grades “C” (scores of 62.7-72.5), “B” (scores of 72.6-78.8), and “A” (scores of 78.9-100). As a SUS score of 80 proves an above-average user experience, it has become a common industrial goal. This threshold was therefore used for interpreting our results.

Covariates
Consultations in general practice are attended by patients of different ages and educational as well as cultural backgrounds, who have a different quantities of digital interactions in everyday life. To determine whether these factors have an influence on the personally perceived usability, we surveyed age, sex, media usage, health literacy, and native language. Information about age and sex were part of the app-taken medical history. In addition to the SUS, we asked patients about which digital media tools were available to them in everyday life (possible answers: cell phone/smartphone, computer/laptop/notebook, tablet, television, none, and others; multiple answers were possible and how many hours a day they used digital media (possible answers: 0≤1, 1≤2, 2≤3, 3≤4, or 4 or more hours). We asked three questions concerning health literacy as a proxy for education attainment, given that educational achievement is the central determinant of health literacy [28]. Questions covering the three aspects of finding/accessing, evaluating/appraising, and understanding health-related information and content were derived from the European health literacy survey [29,30] adapted for the German language (HLS-GER 2 [31]). The HLS-GER 2 uses a predefined 4-point Likert scale.

Table 1. Items of the System Usability Scale (SUS) [19].

<table>
<thead>
<tr>
<th>Items</th>
<th>English version of the statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS 1</td>
<td>I think that I would like to use this system frequently</td>
</tr>
<tr>
<td>SUS 2a</td>
<td>I found the system unnecessarily complex</td>
</tr>
<tr>
<td>SUS 3</td>
<td>I thought the system was easy to use</td>
</tr>
<tr>
<td>SUS 4a</td>
<td>I think that I would need the support of a technical person to be able to use this system</td>
</tr>
<tr>
<td>SUS 5</td>
<td>I found the various functions in this system were well integrated</td>
</tr>
<tr>
<td>SUS 6a</td>
<td>I thought that there was too much inconsistency in this system</td>
</tr>
<tr>
<td>SUS 7</td>
<td>I would imagine that most people would learn to use this system very quickly</td>
</tr>
<tr>
<td>SUS 8a</td>
<td>I found the system very cumbersome to use</td>
</tr>
<tr>
<td>SUS 9</td>
<td>I felt very confident using the system</td>
</tr>
<tr>
<td>SUS 10a</td>
<td>I needed to learn a lot of things before I could get going with this system</td>
</tr>
</tbody>
</table>

*The scores of negatively connoted SUS items were inverted when calculating the sum.*
**Statistical Analysis**

Data from the app were saved into a database and subsequently exported to a tab separated format for further analyses. Participants with two or more missing values of the SUS questionnaire were excluded from statistical analysis. In the case of one missing SUS response, we substituted the missing value with a neutral score of 2, as this method has been used with the SUS in previous research [32].

Sociodemographic data are presented as number and percentage of patients for each categorical data point. Mean and SD were utilized for interval or ratio-scaled data, which has become a common industrial goal. Sociodemographic data were compared between participants with SUS scores <80 and ≥80 using the Fisher exact test for 2×2 tables or the Fisher-Freeman-Halton test for categorical variables and the Wilcoxon rank-sum test for continuous variables. A multiple linear regression was conducted using sex, age, native German language, health literacy score, media usage duration per day, sickness level of the participants, and number of stated complaints in the app as independent variables and the SUS score as the dependent variable. Additionally, the individual SUS items were compared according to sex, age (<65 years vs ≥65 years), German native language, and tablet usage with the Wilcoxon rank-sum test. Data are visually presented as boxplots and radar charts. All analyses were carried out using R (4.1.3 under a GNU license) with the packages fmsb [33], psych [34], tidyr [35], dplyr [36], and ggplot2 [37].

**Results**

**Patient Characteristics**

We aimed to include approximately 400 patients for this study. This target was set to be able to form subgroups and to ensure that all types of patient complaints were included in our sample, including those selected on a limited basis. In total, individual data from 397 participants were included, with 5 participants having one missing SUS item. Figure 2 shows the flowchart of included patients and Table 2 shows the patients’ characteristics.

---

**Figure 2.** Flowchart of patient inclusion in the cross-sectional study capturing patients’ perceived usability of the app. SUS: System Usability Scale.
Table 2. Characteristics of the participants of the cross-sectional study capturing patients’ perceived usability of the app (N=397).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>225 (56.7)</td>
</tr>
<tr>
<td>Male</td>
<td>172 (43.3)</td>
</tr>
<tr>
<td><strong>Age (years), median (IQR)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35.0 (25.0)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>152 (38.3)</td>
</tr>
<tr>
<td>30-65</td>
<td>223 (55.4)</td>
</tr>
<tr>
<td>65+</td>
<td>22 (6.3)</td>
</tr>
<tr>
<td><strong>Native language German, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>328 (82.6)</td>
</tr>
<tr>
<td><strong>Devices used regularly</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Smartphone</td>
<td>389 (98.0)</td>
</tr>
<tr>
<td>Tablet</td>
<td>210 (52.9)</td>
</tr>
<tr>
<td>Computer/notebook</td>
<td>310 (78.1)</td>
</tr>
<tr>
<td>Television</td>
<td>296 (74.6)</td>
</tr>
<tr>
<td><strong>Media usage duration per day (hours), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>89 (22.4)</td>
</tr>
<tr>
<td>2-4</td>
<td>174 (43.8)</td>
</tr>
<tr>
<td>&gt;4</td>
<td>134 (33.8)</td>
</tr>
<tr>
<td><strong>Self-assessed health literacy, median (IQR)</strong></td>
<td></td>
</tr>
<tr>
<td>Understanding doctor</td>
<td>2.0 (0.0)</td>
</tr>
<tr>
<td>Search and understand health information</td>
<td>2.0 (1.0)</td>
</tr>
<tr>
<td>Evaluate health information</td>
<td>1.0 (1.0)</td>
</tr>
<tr>
<td><strong>“How sick do you feel?”</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>I don’t feel sick</td>
<td>32 (8.1)</td>
</tr>
<tr>
<td>Just a little</td>
<td>70 (17.6)</td>
</tr>
<tr>
<td>Fairly</td>
<td>226 (56.9)</td>
</tr>
<tr>
<td>Very</td>
<td>61 (15.4)</td>
</tr>
<tr>
<td>Unbearably</td>
<td>6 (1.5)</td>
</tr>
</tbody>
</table>

*a* Multiple selection possible.

*b* Measured on a 4-point (0-3) Likert-scale (higher scores indicate higher health literacy levels).

*c* Perceived severity of acute complaint.

**Usability for All Participants**

We found a mean total SUS score of 77.8 points, with 54.4% (216/397) of participants having SUS scores of 80 points or higher, indicating high usability of the app overall. Figure 3 shows boxplots of the individual items in which the scores were calculated for each statement. Irrespective of a positive or negative connotation, a higher score indicates a better result. The maximum score that can be achieved for each item is 10.
Usability Stratified by Sociodemographic Factors
We divided the sample into two groups with the cutoff at a SUS score of 80. Participants with a SUS score of at least 80 were significantly younger, reported higher levels of technology device usage, and higher levels of self-assessed health literacy compared to participants with a SUS score below 80 (Table 3).
Table 3. Sociodemographic variables of study participants stratified by System Usability Scale (SUS) score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>SUS score&lt;80 (n=181)</th>
<th>SUS score≥80 (n=216)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>94 (51.9)</td>
<td>131 (60.6)</td>
<td>.09</td>
</tr>
<tr>
<td>Male</td>
<td>87 (48.1)</td>
<td>85 (39.4)</td>
<td></td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>38.0 (26.0)</td>
<td>32.5 (22.0)</td>
<td>.002</td>
</tr>
<tr>
<td>Age group (years), n (%)</td>
<td></td>
<td></td>
<td>.003</td>
</tr>
<tr>
<td>&lt;30</td>
<td>55 (30.3)</td>
<td>97 (44.9)</td>
<td></td>
</tr>
<tr>
<td>30-65</td>
<td>111 (60.2)</td>
<td>112 (51.4)</td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>15 (9.4)</td>
<td>7 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Native language German, n (%)</td>
<td>142 (78.5)</td>
<td>186 (86.1)</td>
<td>.05</td>
</tr>
<tr>
<td>Devices used regularly d, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smartphone</td>
<td>174 (96.1)</td>
<td>215 (99.5)</td>
<td>.03</td>
</tr>
<tr>
<td>Tablet</td>
<td>84 (46.4)</td>
<td>126 (58.3)</td>
<td>.02</td>
</tr>
<tr>
<td>Computer/notebook</td>
<td>136 (75.1)</td>
<td>174 (80.6)</td>
<td>.22</td>
</tr>
<tr>
<td>Television</td>
<td>125 (69.1)</td>
<td>171 (79.2)</td>
<td>.03</td>
</tr>
<tr>
<td>Media usage duration per day (hours), n (%)</td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>&lt;2</td>
<td>48 (26.5)</td>
<td>41 (19.0)</td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>69 (38.1)</td>
<td>105 (48.6)</td>
<td></td>
</tr>
<tr>
<td>&gt;4</td>
<td>64 (35.4)</td>
<td>70 (32.4)</td>
<td></td>
</tr>
<tr>
<td>Self-assessed health literacy, median (IQR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding doctor</td>
<td>2.0 (0.0)</td>
<td>2.0 (1.0)</td>
<td>.16</td>
</tr>
<tr>
<td>Search and understand health information</td>
<td>2.0 (0.0)</td>
<td>2.0 (1.0)</td>
<td>.01</td>
</tr>
<tr>
<td>Assess confidence of health information</td>
<td>1.0 (1.0)</td>
<td>1.0 (1.0)</td>
<td>.19</td>
</tr>
<tr>
<td>&quot;How sick do you feel?&quot;, n (%)</td>
<td></td>
<td></td>
<td>.35</td>
</tr>
<tr>
<td>I don’t feel sick</td>
<td>11 (6.1)</td>
<td>21 (9.7)</td>
<td></td>
</tr>
<tr>
<td>Just a little</td>
<td>30 (16.6)</td>
<td>40 (18.5)</td>
<td></td>
</tr>
<tr>
<td>Fairly</td>
<td>102 (56.4)</td>
<td>124 (57.4)</td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>34 (18.8)</td>
<td>27 (12.5)</td>
<td></td>
</tr>
<tr>
<td>Unbearably</td>
<td>3 (1.7)</td>
<td>3 (1.4)</td>
<td></td>
</tr>
</tbody>
</table>

aFisher exact test.
bWilcoxon rank-sum test.
cFisher-Freeman-Halton test.
dMultiple selection possible.
ePerceived severity of acute complaint.

A multiple linear regression predicting the SUS score was conducted, including sex, age, native German language, health literacy score, media usage duration per day, sickness level of the participants, and number of stated complaints in the app as independent variables (see Table 4). Age, sex, health literacy score, and German native language were significantly associated with SUS score. A higher age ($t_{385}=3.30, P=.001$) and male sex ($t_{385}=1.98, P=.05$) were negatively associated with SUS score, whereby a higher health literacy score ($t_{385}=2.83, P=.01$) and German as a native language ($t_{385}=2.51, P=.01$) were positively associated with SUS score.
Table 4. Multiple linear regression predicting the System Usability Scale sum score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>β (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex (reference=female)</td>
<td>−3.21 (−6.40 to −0.02)</td>
<td>.05</td>
</tr>
<tr>
<td>age (per year)</td>
<td>−.17 (−.27 to −.07)</td>
<td>.001</td>
</tr>
<tr>
<td>German not native language (reference=yes)</td>
<td>−5.39 (−9.61 to −1.17)</td>
<td>.01</td>
</tr>
<tr>
<td>Does not use tablet (reference=yes)</td>
<td>−1.44 (−4.66 to 1.79)</td>
<td>.38</td>
</tr>
<tr>
<td><strong>Average daily media usage (reference=&lt;2 h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4 h</td>
<td>3.21 (−.95 to 7.37)</td>
<td>.13</td>
</tr>
<tr>
<td>&gt;4 h</td>
<td>.01 (−4.46 to 4.48)</td>
<td>.99</td>
</tr>
<tr>
<td>Health literacy score (scale 0-9)a</td>
<td>1.48 (.45 to 2.51)</td>
<td>.01</td>
</tr>
<tr>
<td>How sick do you feel? (score 1-5)b</td>
<td>−1.05 (−2.92 to .82)</td>
<td>.27</td>
</tr>
<tr>
<td>Number of stated complaints (1-11)</td>
<td>−.98 (−2.11 to .15)</td>
<td>.09</td>
</tr>
</tbody>
</table>

aHigher scores indicate a higher level of health literacy.  
bPerceived severity of acute complaint; higher values indicate a higher level of discomfort.

Differences in Individual Items of the SUS

Stratified according to sex, age, native language, and tablet usage (see Figure 4), significant differences were detected in SUS items 2 (“unnecessarily complex”), 4 (“need technical support”), 7 (“learn to use quickly”), 8 (“cumbersome to use”), and 10 (“needed to learn a lot”).

In comparing female and male respondents, all statements were rated more positively by female participants, except for items 1 (“would use frequently”) and 4 (“need technical support”). Female participants also scored significantly higher than male participants on items 2 (“unnecessarily complex”) (mean 7.82 vs 7.11; \( P=.04 \)), 7 (“learn to use quickly”) (mean 8.11 vs 7.53; \( P=.02 \)), 8 (“cumbersome to use”) (mean 8.57 vs 8.08; \( P=.05 \)), and 10 (“needed to learn a lot”) (mean 9.14 vs 8.63; \( P=.03 \)).

Respondents aged 65 years and older scored significantly higher on items 2 (“unnecessarily complex”) (mean 7.58 vs 6.50; \( P=.04 \)), 4 (“need technical support”) (mean 8.72 vs 6.20; \( P<.001 \)), and 10 (“needed to learn a lot”) (mean 9.05 vs 7.; \( P<.001 \)) compared to their counterparts.

German native language speakers scored significantly higher on items 4 (“need technical support”) (mean 8.73 vs 7.75; \( P=.001 \)), 8 (“cumbersome to use”) (mean 8.62 vs 7.10; \( P<.001 \)), and 10 (“needed to learn a lot”) (mean 9.11 vs 8.01; \( P<.001 \)) relative to nonnative speakers.

Lastly, patients who regularly use a tablet had significantly higher SUS scores on items 4 (“need technical support”) (mean 8.95 vs 8.11; \( P<.001 \)) and 10 (“needed to learn a lot”) (mean 9.18 vs 8.64; \( P=.02 \)) in comparison to those of participants who reported reduced levels of tablet use.
Discussion

Principal Findings

In this study, we evaluated the usability of an app in taking medical histories in general practice directly from patients using the SUS [19].

The app achieved a mean SUS score of 77.9, which corresponds to a B+ grade on the curved grading scale [27] and represents a “better” product that does not necessarily need improvement [25]. Other medical devices, even those widely used at home, have lower SUS scores. Kortum and Peres [38] assessed the usability of home health care devices among students, thus representing relatively young, healthy, and well-educated participants. SUS scores for these devices ranged from 65 for
an epinephrine injector to 67 for a pregnancy test kit and 81 for a thermometer, even with previous experience using these devices.

To ensure patients can be active participants in the digital medical history-taking process, the app must be easy and intuitive to use without technical introduction or support. This importance is reflected in items 3 (“easy to use”), 4 (“need technical support”), 7 (“learn to use quickly”), and 10 (“needed to learn a lot”). Mean values between 7.8 and 8.9 for these items indicate that intuitive use has been successfully addressed in the development of our app. Item 1, assessing the frequency of app usage, scored the lowest (6.2), which can be explained by the app’s implementation solely in a medical setting and not utilized regularly in leisure time. As such, this finding is the least meaningful for our purpose.

In a pilot study by Melms et al [39], a self-completed tablet-based digital questionnaire designed for collecting medical histories in an emergency department was found to score high with respect to perceived usability. The design and content were similar to those of our app; however, their questions were only based on the SUS, which does not allow direct comparison. Other comparable instruments, although also for emergency departments, have been tested for usability in pilot studies using self-developed satisfaction surveys [40,41], a single question, and researcher or staff documentation of a patient’s need for assistance [42]. In these studies, patients were mostly satisfied with the self-administered medical history—taking tools and reported good ease of use. Taken together, these results give hope that it is possible to design a medical history app that is perceived as user-friendly.

Nonetheless, obstacles to implementing a digital tool in general practice settings can be multifaceted. Surprisingly, we found that sex was significantly associated with usability; female participants had significantly higher SUS scores than male participants. The fact that men scored higher than women for item 4 (“need technical support”) suggests that men felt more confident than women with using the app. Previous studies demonstrated that men tend to report overconfidence in their abilities, especially in fields with a male connotation [43], which computer science certainly represents [44]. Therefore, it is unclear whether men really would have needed less help or whether they overestimated themselves in their technical skills.

Our study suggests that older people are more likely to have difficulties with the handling of such an app. This aligns with a study showing that from the retirement age of 65 years, digital media use among the German population begins to decline dramatically [45] and a positive attitude toward digitalization decreases with increasing age [46]. Older age has a negative impact on the broad usability score given to a user interface [25]. To that end, this study cannot definitively conclude if the older participants of this study actually perceived the app to be of relatively low usability or if their more negative attitude toward the benefit of new technologies prompted them to give lower scores. Due to the small sample size of participants aged 65 years and older, it is not possible to assert how older people in general would cope with the handling of the app. Since GPs are consulted predominantly by older people [47], further research should focus on app testing with older patients to obtain specific feedback, including suggestions for improvement.

Having learned German as a native language was positively associated with a higher SUS score, although only patients with sufficient German language proficiency were included in the study [17]. This could be due to two different reasons: despite the app’s plain language, it is possible that some of the medical history questions or SUS items were not understood properly.

Daily media use was not associated with the SUS score, which suggests that the app is designed to also ensure that people with limited digital experience do not feel overstrained with its operation.

Limitations

Despite our efforts, this study comes with several limitations. The number of older participants (ie, aged 65 years and above) was relatively low in comparison to their constituents in GP practice settings [47]. One potential reason could be a more pronounced skepticism toward digital tools in older generations, leading to an increase in refusal for participation in the study among older patients. However, as no screening lists were maintained, this is mere speculation. A screening list should be obtained in future studies to be able to characterize individuals who declined participation. Another consideration is that people with lower levels of digital media literacy use may not have agreed to participate in the study.

Data collection was performed during the SARS-CoV-2 pandemic, which may have disproportionately impacted study participants as certain patient groups may have avoided seeing a doctor or were more likely to refuse to participate in the study to avoid unnecessary contact. This could have included especially vulnerable groups such as older people or those with multimorbid conditions.

The Likert scale of the SUS questions shown with clickable singular dots was replaced by a slider on December 8, 2021. In the dot-based representation, it was compulsory to make an entry before continuing, whereas the slider was automatically set to the neutral center and could be shifted. This may have led to incorrectly rated items. For example, this may have occurred in instances of the internet faltering or the patient having double-clicked without noticing. Since there were repeated questions about the word “Inkonsistenzen” (SUS item 2), we replaced it by the more common synonym “Unstimmigkeiten” (English translation: discrepancies). Furthermore, hardware as well as the operating system may have influenced the evaluation of the personally perceived usability of a system [48]. For this study, iPad Minis with the iOS operating system were exclusively used. Therefore, possible differences in assessment related to the operating system and hardware are not part of this study.

The SUS is able to classify the usability of a system but is unable to identify specific usability issues nor capture the usability of the system in its entirety. For a more in-depth usability evaluation, different methods could be used (eg, interviews and observations). During data collection, staff were able to observe usability problems. In their observations, multiple-choice, single-choice, and hybrid questions as well as the slider did not
appear to cause any difficulties. In contrast, problems concerning the handling of the app arose when participants were required to input free-text entries (e.g., age, height, weight). Further, some study participants were unclear on how to open and close the on-screen keyboard. Some participants also did not understand that the figure on which a pain or an injury could be assigned to a body region (see Figure 1E) could be rotated by clicking on an icon at the bottom left of the screen. This means that, for example, back pain may have been falsely reported as abdominal pain. Lastly, an unstable internet connection arose during data collection, which caused the app to be unresponsive intermittently. These factors may have influenced the SUS score.

Conclusion

The app examined in this study for medical history taking passes the usability test based on the SUS and appears to function on par with other digital tools that have become well-integrated in our everyday lives. However, not all people adapted equally well to the app. For successful implementation, all end users, regardless of age, technical affinity, health literacy, or preferred language, must be able to use such a tool. Only if that is attained, providing practical digital solutions can contribute to the efficient and effective delivery of health care services. Therefore, further research should focus on the identification of causes for difficulties of using the app as well as finding appropriate solutions.

Acknowledgments

We would like to thank Julie Ngo, MD for proofreading. We sincerely thank all patients for their participation. We are very grateful to the physicians and their teams who allowed data collection in their practices despite the SARS-CoV-2 pandemic and the extraordinary burden that came along for health care workers. We thank the team at aidminutes GmbH for their commitment and support. This research is funded by the German Innovation Fund (funding number 01VSF19050) of the Federal Joint Committee (G-BA). The funders have not influenced the design of this study and did not play any role during its implementation (e.g., data collection and analysis, interpretation, and publication of the results).

Data Availability

The data sets used and analyzed during the study are not publicly available due to the decision of the research ethics board but can be obtained from the authors upon reasonable request within a data sharing agreement.

Authors' Contributions

KA wrote the original draft. KA and EMN were mainly responsible for writing the manuscript. KA, EMN, and DS conceived the study design and analyzed the data. FM and CJ revised the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Original and German versions of the System Usability Scale (SUS) and the items customized for this study. [DOCX File, 15 KB - humanfactors_v11i1e47755_app1.docx ]

References


33. Package 'fsmb'. CRAN R project. 2023. URL: https://cran.r-project.org/web/packages/fsmb/fsmb.pdf [accessed 2023-12-29]


Abbreviations

- **DASI**: digitally assisted information acquisition before medical consultation
- **GP**: general practitioner
- **HLS-GER 2**: European health literacy survey adapted for the German language
- **SUS**: System Usability Scale

Edited by A Kashniruk; submitted 31.03.23; peer-reviewed by A Simmenroth, R Marshall; comments to author 27.08.23; revised version received 10.10.23; accepted 20.11.23; published 05.01.24.

Please cite as:

Albrink K, Schröder D, Joos C, Müller F, Noack EM

Usability of an App for Medical History Taking in General Practice From the Patients’ Perspective: Cross-Sectional Study

JMIR Hum Factors 2024;11:e47755

URL: https://humanfactors.jmir.org/2024/1/e47755

doi:10.2196/humanfactors.4570

PMID:34905785
A Novel Continuous Real-Time Vital Signs Viewer for Intensive Care Units: Design and Evaluation Study

Shiming Yang1, PhD; Samuel Galvagno1, MD, PhD; Neeraj Badjatia2, MD; Deborah Stein3, MD; William Teeter4, MD; Thomas Scalea3, MD; Stacy Shackelford5, MD; Raymond Fang6, MD; Catriona Miller3, PhD; Peter Hu1, PhD; VS viewer study group7

1Department of Anesthesiology, University of Maryland School of Medicine, Baltimore, MD, United States
2Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, United States
3Department of Surgery, University of Maryland School of Medicine, Baltimore, MD, United States
4Emergency Medicine, University of Maryland School of Medicine, Baltimore, MD, United States
5United States Air Force Academy, Colorado Springs, CO, United States
6Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States
7See Acknowledgments

Corresponding Author:
Shiming Yang, PhD
Department of Anesthesiology
University of Maryland School of Medicine
11 S Paca St, LL01
Baltimore, MD, 21201
United States
Phone: 1 4103284179
Email: syang@som.umaryland.edu

Abstract

Background: Clinicians working in intensive care units (ICUs) are immersed in a cacophony of alarms and a relentless onslaught of data. Within this frenetic environment, clinicians make high-stakes decisions using many data sources and are often oversaturated with information of varying quality. Traditional bedside monitors only depict static vital signs data, and these data are not easily viewable remotely. Clinicians must rely on separate nursing charts—handwritten or electric—to review physiological patterns, including signs of potential clinical deterioration. An automated physiological data viewer has been developed to provide at-a-glance summaries and to assist with prioritizing care for multiple patients who are critically ill.

Objective: This study aims to evaluate a novel vital signs viewer system in a level 1 trauma center by subjectively assessing the viewer’s utility in a high-volume ICU setting.

Methods: ICU attendings were surveyed during morning rounds. Physicians were asked to conduct rounds normally, using data reported from nurse charts and briefs from fellows to inform their clinical decisions. After the physician finished their assessment and plan for the patient, they were asked to complete a questionnaire. Following completion of the questionnaire, the viewer was presented to ICU physicians on a tablet personal computer that displayed the patient’s physiologic data (ie, shock index, blood pressure, heart rate, temperature, respiratory rate, and pulse oximetry), summarized for up to 72 hours. After examining the viewer, ICU physicians completed a postview questionnaire. In both questionnaires, the physicians were asked questions regarding the patient’s stability, status, and need for a higher or lower level of care. A hierarchical clustering analysis was used to group participating ICU physicians and assess their general reception of the viewer.

Results: A total of 908 anonymous surveys were collected from 28 ICU physicians from February 2015 to June 2017. Regarding physicians’ perception of whether the viewer enhanced the ability to assess multiple patients in the ICU, 5% (45/908) strongly agreed, 56.6% (514/908) agreed, 35.3% (321/908) were neutral, 2.9% (26/908) disagreed, and 0.2% (2/908) strongly disagreed.

Conclusions: Morning rounds in a trauma center ICU are conducted in a busy environment with many data sources. This study demonstrates that organized physiologic data and visual assessment can improve situation awareness, assist clinicians with recognizing changes in patient status, and prioritize care.

(JMIR Hum Factors 2024;11:e46030) doi:10.2196/46030
KEYWORDS
clinical decision-making; health information technology; intensive care units; patient care prioritization; physiological monitoring; visualization; vital signs

Introduction
Clinicians working in intensive care units (ICUs) must be able to see, understand, and respond quickly to the complex and ever-changing clinical environment of the ICU. They need to be able to collect, analyze, and interpret what is happening and what it means [1]. Situational awareness is essential for ICU clinicians to provide safe and effective care to their patients. When clinicians have good situational awareness, they are better able to identify and respond to changes in their patients’ condition and to coordinate care with other members of the health care team. However, clinicians are immersed in a cacophony of alarms and a relentless onslaught of data. Within this frenetic environment, clinicians make high-stakes decisions using multiple data sources and are often oversaturated with information of varying quality. While modern hospitals are equipped with bedside monitors collecting various physiological data in a real-time, continuous, and automated way, these data are not always easily accessible remotely or available to be viewed as a continuous trend [2]. The enormous amount of unprocessed data adds an additional burden on ICU clinicians who work in a dynamic environment with voluminous decision-making requirements. Traditional bedside monitors only show a single patient’s instantaneous (static) vital signs (VS) data, limiting the clinician’s scope to view a patient’s physiological trajectory within a clinically meaningful period of time. Clinicians must rely on separate nursing charts—handwritten or electronic—to review a patient’s physiological status. Moreover, auditory alarms often cause “alarm fatigue” instead of increasing situational awareness [3]. Many bedside monitors only display 1 or 2 patients’ information; the ability to view an entire unit or ward allows a clinician to prioritize attention to those in most need of critical care support [4]. Improved visualization of patient information may help clinicians cope with information overload in critical care settings by improving situational awareness and supporting clinical decision-making [5]. An automated physiological data-organizing and information-summary system that presents aggregated information from multiple data sources while providing at-a-glance summaries of clinical data can assist ICU clinicians with prioritizing care for multiple patients.

Developed initially for use in aircraft transporting multiple patients who are critically ill, this VS viewer has 2 outcomes of direct and important clinical applicability. First, the VS viewer can provide clinicians with the capability to monitor individual patient trends, improving overall decision-making. Since patients in the ICU require multiple life support treatments to ensure ideal long-term outcomes, improved display of VS patterns could improve patient assessment and clinical decision-making. Second, the VS viewer system allows remote monitoring of groups of patients through a display that provides clinicians with the ability to quickly identify patients in need of rapid intervention. The objective of this work is to evaluate the use of a VS viewer in ICUs at a high-volume level 1 trauma center. We hypothesized that clinicians would subjectively report improved situational awareness and enhanced ability to make clinical decisions with the use of a VS viewer.

Methods
Data and System Design
In the ICUs of the University of Maryland Medical System, GE Marquette Solar 7000/8000 (General Electric) patient VS monitors are networked to provide a collection of real-time patient VS data streams. Each patient monitor collects real-time 240 Hz waveforms and 0.5 Hz trend data, which are transferred through the secure intranet to a dedicated BedMaster server (Excel Medical Electronics) and archived [6]. To increase the system’s availability and reliability, a triple-redundant design was used, in which 3 BedMaster servers were used in parallel to collect data from all bed units [7]. Physiological data collected through this system, when they are displayed on the GE Marquette monitor, include electrocardiographic, photoplethysmographic, carbon dioxide, arterial blood pressure, and intracranial pressure (ICP), among others. Trends include heart rate (HR), respiratory rate, temperature, oxygen saturation, end-tidal carbon dioxide, and ICP, among many others. This information provides continuous VS data that relays important physiological information regarding brain perfusion, cardiac stability, overall tissue perfusion, and respiratory status.

During the design of the VS viewer for ICU, our goal was to create a novel physiological data displayer that can reduce ICU clinicians’ workload, enhance clinical decision-making, and improve communication in a noisy and confined ICU environment. To achieve the goal, we considered the factors of usability and patient safety, which can be closely related in this application. For usability, current bedside monitors often suffer from insufficient time windows to display physiological trends, a lack of clear indications of patients’ physiological status, and a lack of overview of multiple patients for prioritizing [4]. To enhance the clinicians’ efficiency while maximizing patients’ safety, we adopted the following design strategies: First, the viewer should reduce the information overload for clinicians to access patients’ physiological data, current or past, individual or group [8-10]. Second, it should be compatible with the existing patient monitor system so that clinicians can reuse their existing knowledge about the monitor, which may increase the acceptance of the VS viewer [8]. Third, in the user experience design, the viewer should place the user in control [11]. It should use simple colors and graphs to convey efficient information while still providing detailed data for advanced users to access with simple operations [12]. Fourth, the viewer should have reasonable reliability for patients’ safety. Redundance was introduced in the design for key components in the system, such as the data collection, database, and web server [7].

The VS viewer adopted a client-server architecture. The server handles 2 types of clients: the bedside monitors and the users. It receives and persists in real-time physiologic data that are
transmitted from the bedside monitors. A database records each bedside VS value, bed name, and timestamp. The server also responds to users’ requests for viewing data within a given time frame. To continuously present the latest data to the user with low latency, the VS viewer uses the asynchronous Javascript and XML technique to pull the most recent data from the database every minute [13]. Such a method allows the VS viewer to automatically redraw all VS trajectories without refreshing the entire viewing page.

The VS viewer provides a rich interface for data monitoring, exploration, and recording. Data are depicted according to each clinical area of operations, such as the trauma resuscitation unit or emergency department, operating room, computed tomography suite, and individual critical care units. Figure 1 demonstrates the grouping of bed units. On the left panel, a list of all groups can be used as a shortcut to bed units. Selecting a specific unit, a default 24-hour view is displaced for shock index (SI=HR/systolic blood pressure), HR, systolic blood pressure, ICP, cerebral perfusion pressure, brain trauma index, and end-tidal carbon dioxide concentration. If ICP data are not collected, the space is used to plot the next available VS, optimizing the view.

When a bed is selected, a page for this bed (unit view) is displayed. Figure 2 demonstrates the structure of the information. The page is partitioned into multiple areas for navigation, viewing, and tools. Its center is assigned for presenting the selected patient’s physiologic data in a time frame (up to 72 hours). VS trajectories are stacked vertically in order of predefined importance. The bottom is reserved for plotting bar segments of all VS that summarize the colored warnings without showing the value changes. This provides a summary of all available VS trends in a condensed space, which could be used to view the physiological stability of the patient over time. To provide an at-a-glance view of other rooms in this group, the left panel lists all the rooms in the current group and updates their VS trajectories in real time. The color-coded warning in the thumbnails enhances situation awareness even when the users are focusing on 1 patient.

The VS viewer has additional diagnostic tools. For example, SI is a commonly used blood transfusion diagnosis tool [14]. The VS viewer adds a 2D SI diagram to show a changing trajectory (Figure 3). To present the temporal information, a heat map is plotted, ranging from blue (cold) to red (warm); blue colors represent past events, whereas red colors represent current data trends. Similarly, the brain trauma index (which is ICP or cerebral perfusion pressure) can also be visualized in the 2D plot [15].
Figure 2. Vital sign (VS) viewer in the “unit” mode, with default 24-hour display. Labeled area 1: navigation menu to other room groups. Area 2: title information for room name, current time, and the next update time. Area 3: user portal. Area 4: list of beds in the same group with their current VS thumbnails. Area 5: the main area to display selected room VS trajectories and the summarization with color-coded patterns. Area 6: diagnostic tools for 2D scatter plots of shock index (=heart rate/systolic blood pressure) and brain trauma index (=intracranial pressure/cerebral perfusion pressure). Area 7: functional buttons for selecting various time ranges for viewing.

Figure 3. An example 2D shock index plot. The colored scatter plot shows the change in shock index (heart rate/systolic blood pressure) from past (blue) to recent (red), thereby depicting a 3-day change in worsening shock index.
Clinical Thresholds

Colored warnings are an effective means to gain a clinician’s attention and may be more effective than audible alarms, especially in a noisy, busy, and confined environment [16]. In the VS viewer, VS trajectories with colors may be viewed to highlight the sections where the VS are outside of normal clinical thresholds. For example, too low or too high HR segments are displayed differently from normal HR. Clinical thresholds for VSs were developed after surveying 47 clinicians (24 medical doctors, 18 registered nurses, and 5 respiratory therapists). Among them, 36 clinicians were from the University of Maryland, Baltimore, and 11 from the University of Cincinnati. After the survey was completed, a team of clinicians met to review the results to reach a consensus on the viewer’s opinion of their visual appearance. Multimedia Appendix 1 summarizes the optional threshold distributions for some important VS. Based on these threshold values, a consensus set of color-coded cutoffs was determined (Multimedia Appendix 2). These values were set as fixed parameters under consideration of a simplified and consistent user interface.

Survey Design

Clinicians who were scheduled to work in the ICU or on the trauma teams were contacted and trained on how to use the VS viewer. Once trained, ICU and team clinicians were asked to participate in the study. Clinicians were surveyed anonymously from Tuesday to Friday and were asked to conduct rounds normally, using data reported from nurse charts and briefs from fellows to inform their clinical decisions. None of those clinicians participated in the design of the VS viewer. A total of 2 questionnaires were designed to collect clinicians’ opinions about a patient’s condition and satisfaction with the VS viewer. Clinicians were given a preview survey upon their assessment and formulation of their plan for each patient after traditional rounds and before accessing the viewer. Immediately following the completion of the pre-view survey, the VS viewer was presented to the clinicians on a tablet, displaying the patient’s past physiologic data visualized and summarized for up to 72 hours. After reviewing the viewer for up to 1 minute, clinicians completed the postview questionnaire. In both questionnaires, the clinicians answered questions regarding the patient’s stability, status, and need for a higher or lower level of care. In the post-view questionnaire, clinicians were also asked if they intentionally planned to implement any of the following interventions after seeing the viewer: (1) changing any current medications, (2) ordering additional medications, (3) ordering additional diagnostic tests, (4) changing ventilation settings, (5) ordering additional labs, (6) physically reexamining this patient, (7) providing fluid bolus, or (8) providing a blood transfusion.

Statistical Methods

A participant’s perceiving of the VS viewer’s usefulness is represented by a vector consisting of the percentage of the 5 categories (strongly agree, agree, neutral, disagree, and strongly disagree) that he or she assigned to the question “the viewer enhanced my understanding of the patient’s condition.” We used the Ward method, a hierarchical clustering method, with Manhattan distance to group the participants based on their ratings to the question “the viewer enhanced my understanding of the patient’s condition” [17,18]. Between those clusters, we compared the participants’ opinion changes on the patients’ conditions in 7 questions (Table 1) before and after using the viewer. The chi-square test was used to compare percentage differences.

Table 1. The number of opinion changes for 7 questions (Q1-Q7) before and after seeing the viewer, with respect to the 5 clustered user types.

<table>
<thead>
<tr>
<th>Questionsa</th>
<th>Total changes, n (%)</th>
<th>Unique participants, n (%)</th>
<th>C1, n</th>
<th>C2, n</th>
<th>C3, n</th>
<th>C4, n</th>
<th>C5, n</th>
<th>Like (C1 and C2), n (%)</th>
<th>Dislike (C3, C4, and C5), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>129 (14.2)</td>
<td>16 (66.7)</td>
<td>46</td>
<td>31</td>
<td>10</td>
<td>42</td>
<td>0</td>
<td>77 (59.7)</td>
<td>52 (40.3)</td>
</tr>
<tr>
<td>Q2</td>
<td>112 (12.3)</td>
<td>15 (62.5)</td>
<td>38</td>
<td>34</td>
<td>8</td>
<td>32</td>
<td>0</td>
<td>72 (64.3)</td>
<td>40 (35.7)</td>
</tr>
<tr>
<td>Q3-6</td>
<td>145 (16)</td>
<td>18 (75)</td>
<td>58</td>
<td>54</td>
<td>3</td>
<td>30</td>
<td>0</td>
<td>112 (77.2)</td>
<td>33 (22.8)</td>
</tr>
<tr>
<td>Q7</td>
<td>92 (10.1)</td>
<td>17 (70.8)</td>
<td>20</td>
<td>32</td>
<td>9</td>
<td>31</td>
<td>0</td>
<td>52 (56.5)</td>
<td>40 (43.5)</td>
</tr>
</tbody>
</table>

aPlease refer to Textbox 1 for the question.
Textbox 1. Questions.

- Q1: Having reviewed the last 24 hours of information during rounds and before and after seeing the 24-hour viewer, do they feel that in the past 24 hours the patient has shown evidence of (a) infection, (b) hemodynamic instability, (c) uncontrolled bleeding, or (d) respiratory deterioration?
- Q2: Over the past 24 hours, has the patient’s condition (a) improved significantly, (b) improved slightly, (c) unchanged, (d) deteriorated slightly, or (e) deteriorated significantly?
- Q3: Can the patient be transferred to a lower level of care?
- Q4: Can the patient be transferred to a higher level of care?
- Q5: Does the patient have a traumatic brain injury?
- Q6: Did the patient have intracranial pressure problems in the past 24 hours?
- Q7: Due to the viewer, do they plan for any changes in interventions, including (a) changing any current medications, (b) ordering additional medications, (c) ordering additional diagnostic tests, (d) changing ventilation settings, (e) ordering additional labs, (f) physically reexamining this patient, (g) providing a fluid bolus, or (h) providing a blood transfusion?

Note: These are the questions referenced in Table 1.

Ethical Considerations
The study has been approved by the institutional review board of the University of Maryland School of Medicine (HP-00063086).

Results
Survey Collection
From February 2017 to June 2017, the survey team followed clinicians who agreed to take the surveys. A total of 908 surveys were collected from 24 participants with unbalanced proportions. Among the 908 rounds, 48 (5%) were patients who were newly admitted, and 860 (95%) were not. When asked if the VS viewer enhanced their understanding of the patient’s condition, clinicians strongly agreed 45 (5%) times, agreed 514 (56.6%) times, were neutral 321 (35.4%) times, disagreed 26 (2.9%) times, and strongly disagreed 2 (0.2%) times. Figure 4 lists the total surveys each participant contributed and the proportions of ratings on whether the viewer enhanced his or her understanding of the patient’s condition during a round.

Results show that physicians’ clinical assessments and plans could be influenced by viewing the VS viewer for 1 minute or less, indicated by a “yes” answer to at least 1 of the 8 questions (Q7 in the survey). Of the 908 rounds, a total of 92 (10.1%) rounds had at least 1 “yes” as planning on some changes to the interventions. The most common change was (Q1) changing current medications (36/908, 4%). The next most common changes were (Q6) physically reexamining the patient (31/908, 3.4%), (Q2) ordering additional medications (20/908, 2.2%), and (Q7) providing a fluid bolus (20/908, 2.2%).

We used the Ward method with Manhattan distance to group the participants based on their ratings to the question “the viewer enhanced my understanding of the patient’s condition” [17]. For example, 1 participant contributed 62 surveys and rated 2 “strongly agree,” 22 “agree,” 32 “neutral,” 5 “disagree,” and 1 “strongly disagree.” The vector of percentages (0.03, 0.35, 0.52, 0.08, and 0.02) represents the overall rating that this participant had about the viewer. The 24 participants were clustered into 5 groups, as shown in Figure 5. The 5 groups correspond to the participants who are mostly in favor (C1) of the viewer to those least in favor (C5). There are 6 in C1, 6 in C2, 3 in C3, 7 in C4, and 2 in C5, which shows a very balanced grouping, with half of the participants in the C1 and C2 groups and the other half in the other 3 clusters. This shows that the sampled rounds were done by participants with almost similar proportions of different attitudes toward the viewer. In other words, the survey team sampled the rounds randomly enough so that the collected data were not biased by participants with certain preexisting feelings about the viewer.
Comparisons
We analyzed the opinion changes before and after seeing the viewer, regarding the patient’s stability, status, and need for a higher or lower level of care. Instead of summarizing the total changes in opinions, we compared them with respect to the clusters of user types. The participants who were “neutral” (C3) or “strongly dislike” (C5) had low numbers of opinion changes.
for all 7 questions. Those who were in clusters C1, C2, and C3 had more numbers of opinion changes (Table 1). For simplicity, we can further group the participants into 2 types: those who liked the VS viewer (C1 and C2) and those who disliked it (C3, C4, and C5). The clinicians who liked the VS viewer had a higher rate of changed opinions than those who disliked the VS viewer regarding Q1 to Q6 (Q1: 59.7% vs 40.3%, Q2: 64.3% vs 35.7%, and Q3-6: 77.2% vs 22.8%). When asked if they planned for any changes for interventions (Q7), there was no significant difference between the 2 major groups of clinicians (56.5% vs 43.5%, \(P=.10\)).

**Discussion**

**Principal Results**

With the development of sensor and computing technologies, vast amounts of high-quality, continuous electronic data, including VS, alarms, and clinical interventions, are collected at the bedside. Those data have the potential to provide an unprecedented view of dynamic physiologic responses to injury, illness, and treatments. Therefore, data gathered from bedsides could assist clinicians in care planning and decision support. However, massive amounts of data that are not well organized or presented still create a barrier for clinicians making full use of them in a busy resuscitation or intensive care environment. Bedside monitors often only display instantaneous readings or a short strip of recent physiologic VS for diagnosis. Clinicians need to rely on separate nursing charts, handwritten or electronic, to review a patient’s developing conditions. The VS viewer, which automates physiological data by displaying clear color-coded trends, presents aggregated information from multiple data sources, provides at-a-glance summaries of clinical data, and assists with the prioritization of care for multiple patients.

The use of the VS viewer was subjectively assessed with 908 observations from clinicians working in ICUs at a high-volume level 1 trauma center. Clinicians generally perceived the use of the VS viewer favorably, as evidenced by survey data. The VS viewer was originally developed for the United States Air Force Critical Care Air Transport Teams [19,20]. Critical Care Air Transport Teams transport up to 3 patients who are critically ill in the back of the aircraft, allowing trauma surgeons to perform far-forward damage control surgery, knowing that these patients could be quickly transported rearward with full support. This rapid transport of complex patients with multisystem trauma, shock, burns, and respiratory failure who are in hemodynamic flux requires continual resuscitation, stabilization, advanced care, and life-saving interventions during air transport; however, currently available advanced ICU monitoring systems suitable for the needs of such patients were developed for use in stable, hospital-based settings, not in the crowded, noisy, vibrating, and sometimes frankly jolting environment of air evacuation or long-distance air transport. The noise levels, confined space, limited access to patients, vibration, and overall limited patient visibility make using a VS viewer advantageous in such a setting. Such technology can also be valuable in enhancing emergency medical personnel’s decision-making for initial triage. While traditional VS are useful in guidingprehospital care and triage, they represent isolated points in time, and trends and fluctuations in vitals may not be apparent.

In this study, we set the clinical thresholds for colored warnings to be uniform across all beds. This was to make the user interface simplified and more consistent during a survey. Additionally, a set of predefined thresholds from a group of experienced clinicians could be a useful out-of-the-box feature when the VS viewer is deployed in the field. That said, the clinical thresholds could be personalized for each bed. For example, if the bedside monitor allows alarm threshold settings, such settings could be used as the colored warning thresholds in the VS viewer for each bed.

The VS viewer has expanded from ICUs to trauma resuscitation units, operating rooms, neuro ICUs, and pediatric ICUs at the University of Maryland Medical Center. In 2020, during the COVID-19 pandemic, it was deployed to monitor 150 beds in bioccontaminated units to reduce the risk of infection and improve efficiency for clinicians in treating their patients.

**Innovations**

The VS viewer is a multipatient physiological monitor. To the best of our knowledge, we could not find any articles that describe a viewer system with a similar design. In a comprehensive review by Waller et al [5], a total of 17 information displays in ICU settings were designed for specific disease states or body systems, such as cerebral perfusion monitoring for individual patients or monitoring for arterial blood gas trends. The novel user interface presented in this study was designed with the aim of conveying information more efficiently to ICU clinicians in a noisy, confined, and busy environment. It uses color-coded warnings to indicate a patient’s status and highlight data that needs attention. The side panel provides a peek at the physiological status of other patients, which can help clinicians keep an eye on other patients even if their attention is focused on a single patient. It uses advanced web front-end techniques to hide large quantities of data behind simple line charts and reveal them when needed.

**Clinical Impact**

The use of the VS viewer can have several possible influences on clinical assessment and plans. It can help clinicians quickly recognize critical changes in the patient’s physiologic status and provide early interventions to prevent further deterioration. The VS viewer can potentially improve patient outcomes by providing clinicians with a concise overview of key information, reducing cognitive load and errors, and improving compliance with evidence-based safety guidelines [12]. It may also help to improve communication efficiency within the ICU team by providing easy access to a shared platform of patient longitudinal data. It can reduce the workload of the ICU team by automating routine tasks such as extracting data from nursing charts.

To prioritize care in high-volume ICUs, intensive care clinicians must be able to rapidly identify physiological events and the need for intervention. The VS viewer can help organize a large amount of data in a busy, noisy ICU environment where close monitoring of patients who are critically ill is essential to detect potentially harmful physiological trends. The presentation of data with temporal, color-coded patterns, and the ability of the
VS viewer to provide at-a-glance data for entire units is advantageous for clinicians working in high-volume ICUs.

The color-coded patterns may reduce the “alarm fatigue” issue in noisy ICUs. The noise burden is common in modern physiologic monitoring systems and has been recognized as a critical patient safety concern in the hospital care setting [21-23]. In noisy environments, such as ICUs, helicopter transportation, or aeromedical evacuation, loud and continuous alarms could reduce their specificity in getting clinicians’ responses. Another issue with audible alarms is that they are transient and cannot be replayed once they are gone. While the visual alert patterns could show the longitudinal patterns of physiologic change.

Related Work

The VS viewer with organized and easy-access information could be part of the effort to build the smart ICU or the tele-ICU. The concepts of smart ICU and tele-ICU aim to maximize the use of bedside clinical expertise in assessing and treating patients by providing integrated monitoring and actionable information [24-26]. A survey study of 86 ICU staff in a German university hospital summarized that health providers expect ICU monitoring could be improved by reducing false alarms, using wireless sensors and mobile devices, preparing for the use of AI, and enhancing the digital literacy of ICU staff [27,28]. The VS viewer could be used in both centralized and decentralized architectures of tele-ICU for extending coverage and facilitating patient transfer between hospitals because of its flexible configuration of grouping ICU beds virtually [29]. By making essential clinical information available remotely, the VS viewer allows clinicians to provide care plans when on-site support is infeasible or limited [30,31]. It may potentially reduce exposure to contagious diseases and, hence, increase patient safety.

With continuous physiologic data and other clinical information, the VS viewer has the ability to process real-time data into predictive algorithms, which is also desired for tele-ICU [30]. Beyond being a plain display, the VS viewer could embed risk-prediction algorithms that use continuous VS as inputs and may promote more efficient interventions to reduce ICU risk [31]. For example, ICU mortality prediction [32,33], secondary insults after severe brain traumatic injury [34], needs for transfusion [35,36], and neurologic decline in the ICU [37] are reported to have good predictive performances by using variables derived from continuous VS. We have also shown that using risk scores calculated from continuously measured VS, patients requiring endovascular resuscitative interventions can be identified with high accuracy [38]. Moreover, the VS viewer could serve as a platform for predictive model diagnosis by providing clinicians with explainable artificial intelligence [39]. With patient VS data, we can use the Shapley Additive exPlanations algorithm to calculate each variable’s contribution to the prediction result [40]. Therefore, the clinicians would know not only the prediction but also the contribution of each variable to the prediction. Such information may help clinicians make more personalized care plans.

Limitations

There are limitations to this work that are worth noting. We collected data from a large number of ICU clinicians compared to trauma team clinicians. Trauma team clinicians are surgeons responsible for the same patient throughout the entire length of stay, regardless of the acuity of the patient. ICU clinicians are intensivists and are only responsible for patients in the ICU. Hence, disparities between both groups of clinicians are inevitable, as each group has different clinical perspectives and patient workloads. As occurs in nearly all survey work, response rates and receptiveness to the surveys varied. Some clinicians were more amenable to being surveyed compared to others. In the collected forms, there were more surveys from some clinicians than from others. To reduce this potential bias, we clustered the participants based on their overall rating on each round, from which we estimated each participant’s a priori attitude toward using this viewer. The results show that there was a balanced “favoring” and “non-favoring” of using this viewer.

We only evaluated the viewer based on clinicians’ satisfaction and efficiency (potential changes in interventions before and after seeing the viewer). In future studies, randomized controlled trials can be designed to analyze the viewer’s impact on patients’ outcomes and safety [12].

Conclusions

We designed, implemented, and evaluated an automated physiologic data organizer and visualization platform. It provides at-a-glance summaries and assists with prioritizing care for multiple patients. The VS viewer demonstrates a method to assemble large quantities of data from multiple sources and represents trends in each patient’s condition with simple color codes, greatly improving situational awareness. It has the potential to be used in en route care, hospitals with multiple branches, and understaffed hospitals in remote areas. The survey shows that organized physiologic data and visual assessment could assist clinicians in recognizing changes in patient status and prioritizing care.

Acknowledgments

This work was partially funded by USAF 8650-14-2-6D19 and W81XWH-17-C-0034.

The authors thank the VS viewer study group: Colin Mackenzie, MD, Sarah Wade, BS, Lauren Hartsby, BS, Thomas Grissom, MD, Napoleon Roux, MD, Doug Floccare, MD, Ashton Engdahl, BS, Scott Murray, BS, Rosemary Kozar, MD, Sarah Murthi, MD, Laura Buchanan, MD, Neal Reynolds, MD, Gunjan Parikh, MD, Megan Brenner, MD, without whose help the data collection for this study would not have been possible.
Conflicts of Interest

PH, DS, Colin Mackenzie (part of the VS viewer study group), TS, and SY have US Patent Application 17/676,657 filed on February 21, 2022, titled “Method and Apparatus for Monitoring Collection of Physiological Patient Data.”

Multimedia Appendix 1
Surveyed thresholds for heart rate, systolic and diastolic blood pressure, blood oxygen saturation, and temperature.
[DOCX File, 20 KB - humanfactors_v11i1e46030_app1.docx]

Multimedia Appendix 2
VS Viewer color coding threshold values.
[DOCX File, 18 KB - humanfactors_v11i1e46030_app2.docx]

References


User-Centered Design and Usability of a Culturally Adapted Virtual Survivorship Care App for Chinese Canadian Prostate Cancer Survivors: Qualitative Descriptive Study

Karen Young¹,², BCom, MSc; Ting Xiong¹,², BSc, MSc; Rachel Lee¹, BSc, MHI; Ananya Tina Banerjee³, PhD; Myles Leslie⁴, PhD; Wellam Yu Ko⁵, PhD; Quynh Pham¹,²,⁶,⁷,⁸, PhD

¹Centre for Digital Therapeutics, Techna Institute, University Health Network, Toronto, ON, Canada
²Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
³Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada
⁴School of Public Policy, University of Calgary, Calgary, AB, Canada
⁵Men’s Health Research Program, University of British Columbia, Vancouver, BC, Canada
⁶Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada
⁷Telfer School of Management, University of Ottawa, Ottawa, ON, Canada
⁸School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Quynh Pham, PhD
Centre for Digital Therapeutics
Techna Institute
University Health Network
Toronto General Hospital, R. Fraser Elliott Building, 4th Floor
190 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 (416) 340 4800 ext 4765
Email: q.pham@uhn.ca

Abstract

Background: Cultural adaptations of digital health innovations are a growing field. However, digital health innovations can increase health inequities. While completing exploratory work for the cultural adaptation of the Ned Clinic virtual survivorship app, we identified structural considerations that provided a space to design digitally connected and collective care.

Objective: This study used a community-based participatory research and user-centered design process to develop a cultural adaptation of the Ned Clinic app while designing to intervene in structural inequities.

Methods: The design process included primary data collection and qualitative analysis to explore and distill design principles, an iterative design phase with a multidisciplinary team, and a final evaluation phase with participants throughout the design process as a form of member checking and validation.

Results: Participants indicated that they found the final adapted prototype to be acceptable, appropriate, and feasible for their use. The changes made to adapt the prototype were not specifically culturally Chinese. Instead, we identified ways to strengthen connections between the survivor and their providers; improve accessibility to resources; and honor participants’ desires for relationality, accountability, and care.

Conclusions: We grounded the use of user-centered design to develop a prototype design that supports the acts of caring through digital technology by identifying and designing to resist structures that create health inequities in the lives of this community of survivors. By designing for collective justice, we can provide accessible, feasible, and relational care with digital health through the application of Indigenous and Black feminist ways of being and knowing.

(JMIR Hum Factors 2024;11:e49353) doi:10.2196/49353
Introduction

Digital health has been posited as a pathway to more equitable and holistic care [1,2]. However, the digital divide, or the capacity for digital technology to exacerbate inequities, has been widely described [3]. Its differential impacts on the social determinants of health are known as the digital determinants of health [4]. Recent years have seen an acceleration of digital health innovations (DHIs) such as digital therapeutics into health care systems, which was supercharged by the COVID-19 pandemic and the resulting widespread implementation of teledmedicine [4]. One such digital therapeutic is the Ned Clinic (“No Evidence of Disease”), which aims to optimize clinical care and patient self-management through virtual asynchronous care delivery for prostate cancer (PCa) survivors [5]. The Ned Clinic platforms, including clinician-led (Specialist Ned) and nurse-led (Ned Nurse) interventions, were developed at the University Health Network in Toronto, Canada, by a consortium of stakeholders [5].

PCa is the most commonly diagnosed nonskin cancer for Canadian male individuals, and most (99%) are estimated to be diagnosed in male individuals aged 50 years and older [6]. Older adults are negatively impacted by the digital divide [7]. Race, a social determinant of health, is also linked to worse survivorship and care outcomes for PCa survivors, most notably for Black male individuals [8]. Asian (generally defined as East Asian and South Asian ethnicity) male individuals have been found to have better survival rates than the median but are more likely to present with advanced PCa, suggesting systemic issues with identifying health issues and obtaining timely appropriate care [9]. These differences carry over into the delivery of follow-up care, as PCa survivors’ care needs and access to care are affected by the complex intersection of ethnicity, culture, and other social and structural factors [10,11].

Cultural adaptation is the process of applying changes to existing health interventions based on “surface” (social and behavioral characteristics) and “deep” (worldview, norms, beliefs, and values) cultural structures [12]. As these structures are known to impact beliefs about illness and well-being, the intent is to provide intervention benefits for communities that have experienced health inequities [13]. Culturally adapted DHIs appear to have been most widely reported in the field of mental health; in contrast, cultural adaptations of cancer survivorship apps have not been published, likely owing to the few determinants of health. These framings can serve to “museumize” and problematize identity categories and culture as causes of ill-health, echoing the long-standing use of culture as a scapegoat to fault specific communities for health inequities. Moreover, defining “culture” for such adaptations can be a complex process in Canada, where culture, race, ethnicity, settler colonialism, and white supremacy (ie, the social and structural determinants of health) all create intersectional and differential lived experiences under a putatively shared identity—Canadian [19-21].

This research reports on the second and final phase of a project to design a cultural adaptation of the patient-facing Ned Clinic virtual follow-up care app for Chinese Canadian PCa survivors. In phase 1, we completed formative work distilling a set of themes relevant to survivors’ user needs for follow-up and virtual care. Following the user-centered design (UCD) framework, we describe the results of the design and formative evaluation of a culturally adapted prototype of the app.

Methods

Study Design

The overall qualitative descriptive study design was structured using the community-based participatory research (CBPR) and UCD frameworks [22-24]. This study was conducted at the University of Toronto between December 2022 and March 2023 during the COVID-19 pandemic. For communities that face barriers to care, it was found that CBPR practices such as our engagement of a key informant and invitations to community members to share their lived experiences through open-ended interviews are appropriate [1,25]. CBPR concepts were applied to meaningfully involve the community (including several authors of this study) and return the results for their benefit. Here, community represents a “symbolic totality as well as a practical multiplicity,” as the Chinese Canadian community is highly heterogeneous [26]. We view our participants as a coalition of self-identified Chinese Canadian individuals impacted by PCa survivorship to attend to their differences.

The Chinese Canadian community is an immigrant community that exists as a result of settler colonialism. In recognizing this, we redefine “immigrants” as “people with ancestral roots outside of Indigenous lands, who are beholden to Indigenous laws and epistemologies” [27]. This definition led us to apply a relational paradigm to this project and an axiology of relational accountability. It also provided a pathway to apply several multilevel Indigenous and Black feminist theorizations, guiding principles, and tools [27-29]. These included decolonial theory, Etuaptmunk (two-eyed seeing), intersectionality, and cultural safety to inform our conceptualization of digital space as intimately related to land [27,30,31]. This approach allowed us to contextualize the place-related experiences of our participants and uncover their desires for relational and culturally safe care [32]. We noted that these desires are not specifically Chinese,
and this presented an opportunity to design for relationally connected digital health.

UCD is a flexible, iterative, and evidence-based 3-step design process framework that consults, involves, and considers the needs of the end user throughout the entire project [23]. Phase 1 of this study encompasses steps 1 and 2; phase 2 encompasses steps 2 and 3. We present this study according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [33].

Step 1: Ideation and Concept Generation
To contextualize the potential use of this app, we sought to understand the structures that impact Chinese Canadian PCa survivors’ experiences with follow-up care and virtual care. The results of this phenomenologically informed exploratory-descriptive qualitative study are described elsewhere [34]. Based on the findings of this formative research, we synthesized a list of design principles (Table 1), which we then categorized into the cultural adaptation taxonomy created by Spanhel et al [14] to systematically adapt the patient-facing prototype.

<table>
<thead>
<tr>
<th>Research finding</th>
<th>Design principle</th>
<th>Taxonomic classification [14]a</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHIb freedom: patients felt that they were expected to track and remember overwhelming amounts of information.</td>
<td>The system should automatically update, store, and provide access to PHI on demand.</td>
<td>• (9) Goals of treatment • (10) Methods of treatment</td>
</tr>
<tr>
<td>Access to personalized education and information: patients felt that they were unable to access information about their care options and disease status.</td>
<td>The system should provide access to personalized and evidence-based information regarding staging, self-management, and treatment options.</td>
<td>• (9) Goals of treatment • (10) Methods of treatment</td>
</tr>
<tr>
<td>Continuity of care: patients desired a connection with their provider and the ability to communicate during times of need.</td>
<td>The system should improve accessibility and continuity of care, as strong care relationships create a sense of safety.</td>
<td>• (9) Goals of treatment • (10) Methods of treatment</td>
</tr>
<tr>
<td>Security: patients expressed suspicion about digital health because they had concerns about surveillance and security.</td>
<td>The system should be architected and built with a high level of security and privacy.</td>
<td>• (12) Functionality</td>
</tr>
<tr>
<td>Accessibility: patients wanted to access care in readable and accessible language formats.</td>
<td>The system should provide readable language and accessible language formats.</td>
<td>• (5) Language translation • (6) Language tailoring</td>
</tr>
<tr>
<td>Digital literacy: patients felt comfortable with their device of choice but desired simplicity, form over function, and accessible help and documentation.</td>
<td>The system should prioritize usability, provide straightforward instruction and support, and maintain simple user interface and user experience design.</td>
<td>• (11) Structure • (12) Functionality • (13) Design and aesthetics</td>
</tr>
<tr>
<td>Care coordination: patients felt like they were expected to coordinate their care, as communication between specialists, primary care, and other services were fragmented.</td>
<td>The system should coordinate and provide a clear follow-up appointment schedule.</td>
<td>• (9) Goals of treatment • (10) Methods of treatment</td>
</tr>
<tr>
<td>Resources: patients felt unable to access, refused, or unaware of needed resources such as mental health support.</td>
<td>The system should provide accessible pathways to resources, such as psychological support, supportive care, and financial support.</td>
<td>• (8) Difference in concepts of mental health and its treatment • (9) Goals of treatment • (10) Methods of treatment</td>
</tr>
</tbody>
</table>

aThe design principles used to adapt the Ned Clinic patient app identified here are classified to the corresponding taxonomic components found in Spanhel et al [14].

bPHI: personal health information.

Step 2: Design and Development
We applied these design principles to adapt the Ned Nurse patient app for Chinese Canadian survivors. A composite profile of a sample representative user was created to situate the design team during the development of the wireframes. A list of 5 use scenarios was created to guide the adaptation. These scenarios encompassed the design principles created in step 1 and included actions such as completing follow-up tasks, accessing a follow-up care schedule, and using the app to chat with a clinician. All use scenarios are described in the interview guide (Multimedia Appendix 1). Then, the original Ned Nurse app wireframes were redesigned to reflect the features required to operationalize these scenarios through the app, resulting in a new prototype. The prototype was created in Figma (Figma Inc.) on an iPhone 13 (Apple Inc.) interface. This initial adaptation was iteratively critiqued by a team of researchers and human factors designers to refine the content, user interface, and user experience. Once the adapted prototype was finalized, it was translated from English into written Chinese via the
translation process outlined in Haldane et al [35]. This resulted in 3 versions of the adapted prototype in English, Simplified Chinese, and Traditional Chinese. The app homepage in each language version is shown in Figure 1.

Figure 1. Wireframes of the adapted Ned Nurse homepage in all 3 language versions.

Step 3: Evaluation

Overview

We empirically evaluated the acceptability, appropriateness, and feasibility of the adaptation through a cultural safety lens [22]. These dimensions are early-stage implementation outcomes and have also been found to be core to the success of DHIs [36,37]. A moderated cognitive walkthrough approach and the think-aloud protocol were used to construct a semistructured interview guide encompassing the 5 scenarios describing usual tasks that an end user might complete through the app [38].

Usability Testing

Facilitators began each test by outlining the usability testing procedure and think-aloud protocol. Context regarding the intended use and deployment of the Ned Nurse system was provided. Participants were asked to complete a series of actions for each scenario on the prototype to evaluate its design and functionality. We asked participants to think and speak about improvements they desired during their evaluations. In situations where the participant was unable to access the prototype on their device, they were asked to state their intended actions using the think-aloud protocol to the facilitator, who completed the action in the prototype on their behalf.

Interviews were completed through Microsoft Teams or Zoom (Zoom Technologies Inc). Informed consent for this work was previously obtained as part of overall study consent from participants. Participants were provided with the choice of completing their interview in Cantonese, Mandarin, or English and were also able to choose which language they wished to test the prototype in. The results of each usability test were iteratively analyzed via content analysis. Audio recordings of the participant interviews were translated into English as needed, according to the translation process described previously. A deductive and inductive content analysis approach was used, in which analysis of the data was completed by coders (TX and KY) through a process including open coding, creating categories, and abstraction [39]. Recommendations were applied in real time to create a final prototype that incorporated feedback from each user over the course of usability testing.

Positionality

An important marker of excellent qualitative research is “sincerity” or positionality, which indicates that the researcher has thought about and is reflective and aware about their values,
experiences, biases, and inclinations within their research [40]. Here, the lead researcher reports on their social position, personal experiences, and political and professional beliefs to center the active role that the researcher plays in the framing of the research problem, interpretation of data, methods used, and the reporting of the results [41].

KY is a health informatics trainee and second-generation Chinese Canadian settler who was born and raised in the Greater Vancouver Regional District (GVRD) by a working-class, first-generation immigrant family with roots in southeastern China. She does not have any direct experience with PCa and has not previously provided care for PCa survivors. KY works primarily from a relational paradigm, focusing on the structures, contexts, and relationships that shape the design, development, and implementation of digital therapeutics and health technologies. She led and participated in all study activities.

Setting and Place
This study was conducted in the GVRD, located on the current, unceded, and future territories of the (Tsleil-Waututh, Squamish, and Musqueam) First Nations. The GVRD is home to one of Canada’s oldest and largest living Chinese communities, including persons and families whose stories and identities span multiple geographies and generations [42]. The lead (KY) and senior author (QP) established relations with a supportive care program that provides care for Chinese Canadian PCa survivors and a Chinese PCa support group in this area. A key community informant agreed to guide this study and review and approve study materials.

Ethical Considerations
Research ethics approval for this study was obtained from the University of Toronto research ethics board (Human Protocol #43145). Written and verbal informed consent to participate in both phases of the project was obtained from all participants prior to interviews via the REDCap tool (Research Electronic Data Capture; Vanderbilt University), hosted at the University of Toronto. All data collected and disseminated here have been de-identified. Participants were provided with an honorarium of $50.00 CAD ($37.65 USD) per hour in appreciation of their time.

Results
Demographics
Usability testing was performed by 6 user testers, convenience sampled from the pool of 14 survivors and partner-caregivers who participated in the first phase of work as a form of member checking. This sample was also informed by Nielsen-Norman usability testing guidelines [43]. The reasons for nonparticipation were not collected. To protect the privacy of the participants involved in this phase, a demographic overview of the overall research project is provided here. Of the 14 participants in the first phase of this project, all survivors identified as men (n=12, 86%), and all partner-caregivers identified as women (n=2, 14%). A total of 13 (93%) participants indicated that they spoke English as an additional language. Most made an income between CAD $15,000 (US $11,048) and CAD $100,000 (US $73,653; n=12, 86%), lived in an urban area (n=13, 93%), were married (n=12, 86%), were educated beyond high school (n=13, 93%), and were retired (n=9, 64%). A 50/50 split emerged between preferences for smartphone or desktop or laptop use. Most (n=10, 71%) self-rated as being comfortable with their device. Participants indicated that they had 2 or fewer smartphone health apps (n=13, 93%).

Phase 1: Ideation
Table 1 summarizes the user requirement findings that emerged from previous formative research in phase 1 of this project and their subsequent translation to design principles.

Phase 2: Design and Development
Overview of Ned Nurse
An overview of the Ned Nurse clinical trial protocol is described by Pham et al [5]. The findings from formative work on the perspectives of health care providers, patients from the wider PCa survivor community, and the service design of the platform are forthcoming. Briefly, Ned Nurse digitally operationalizes a nurse-led model of survivorship care. Patients complete a series of tasks or access resources designed to support them in their survivorship. The platform aims to facilitate holistic care for patient quality of life.

Overview of the Adapted Patient-Facing System
The patient-facing adaptation set 2 user-input “care tasks,” a validated questionnaire (Expanded Prostate Cancer Index Composite-Clinical Practice [EPIC-CP]) and a needs assessment survey, to constitute a single Ned Nurse “review” [5,44]. Language within the app avoided wording such as appointment, visit, and so forth to clarify the differences between synchronous and asynchronous care encounters. The user interface and user experience were designed to draw the user’s attention to these tasks on the homepage immediately after login. All features were accessible via an in-app hamburger menu.

User inputs to the questionnaire were triaged via a decision-tree algorithm [45]. The algorithm was designed to return in-app self-management resources within a progress note (“Nurse’s Note”) automatically available to the user after input submission. If the algorithm detected that the patient required further support, they were prompted to specify domains for follow-up and asked to select their preferred contact method. This action would flag this patient to the nurse for follow-up. Resource links would appear on the homepage after the note was read and cleared.

To ensure that patients were aware of their review schedule, a feature was designed to display the last date, frequency, and next date of their expected reviews. The name of the nurse in charge and an explanation of their Ned Nurse role were provided to strengthen the perceived connection between the user and the nurse. This feature also set expectations for manual response times and included a link to users’ previous submissions for on-demand access.

Resources were made available in 3 separate categories: symptom self-management advice, PCa information and education, and support and programmatic resources. Within
each category, resources were further categorized. For example, symptom management included resources for symptoms such as anxiety, urinary incontinence, and hot flashes. Each resource provided an overview: relevant self-management steps; off-app links; and the ability to email, print, or save the resource. The feature home page also sectioned resources saved by the user (“Saved Resources”) and resources picked for the user (“Picked for Me”) by their nurse.

All available and historical prostate-specific antigen and testosterone blood work results were made available in chronological order to the user on-demand in a separate feature. Finally, a chat feature was designed to explore whether users might find it useful. It incorporated both responses in English from an automated support assistant (chatbot) and manually submitted by the nurse. This feature was simulated for evaluation.

Phase 3: Evaluation
Of the 6 participants, 2 (33%) tested in Cantonese, 3 (50%) tested in English, and 1 (17%) tested in Mandarin. These ratios correspond to testing of the Traditional Chinese, English, and Simplified Chinese versions. We note that patients who completed their testing in 1 language were functional to fluent in 1 or all of the other languages and provided critique for multiple versions.

Overall, there was strong agreement that the adaptation presented here would be acceptable, appropriate, and feasible for use, with the exception of the chat feature. Participants agreed that this app would make them feel comfortable and safe by allowing them to have more control over their care, access to resources, and stronger connections to their providers. They were encouraged by its perceived ability to meet their needs by protecting their connection with their providers, leveraging the functional flexibility of digital health, and providing resources beyond what they currently accessed. It was particularly valuable that features could be accessed at their convenience, as some felt that their follow-ups were far too short to meet their needs. Overall, 5 (83%) of 6 participants indicated that the level of support provided by this app was beneficial enough that it should be offered to patients prior to beginning treatment, or even at the point of diagnosis.

Participants’ critiques centered on expanding flexibility, access to information, and streamlining responses. They felt that responses for some assessment questions (from 4 to 8 options) were overwhelming and should be reduced (3/6, 50%). English-Chinese translations would increase self-confidence in navigating the health care system. Medication names were encouraged by its perceived ability to meet their needs by allowing them to have more control over their care, access to resources, and stronger connections to their providers. They were encouraged by its perceived ability to meet their needs by protecting their connection with their providers, leveraging the functional flexibility of digital health, and providing resources beyond what they currently accessed. It was particularly valuable that features could be accessed at their convenience, as some felt that their follow-ups were far too short to meet their needs. Overall, 5 (83%) of 6 participants indicated that the level of support provided by this app was beneficial enough that it should be offered to patients prior to beginning treatment, or even at the point of diagnosis.

Participants’ critiques centered on expanding flexibility, access to information, and streamlining responses. They felt that responses for some assessment questions (from 4 to 8 options) were overwhelming and should be reduced (3/6, 50%). English-Chinese translations would increase self-confidence in navigating the health care system. Medication names were highlighted as confusing by some because the connection between hormonal function and fatigue was not readily apparent. The spiritual domain in the needs assessment was flagged, as some thought that it would not be appropriately addressed by the nurse. Those who felt uncomfortable with this domain noted that they would prefer speaking about these needs to a spiritual leader. Agreement on appropriate response times also varied.

The chat function was deemed possibly helpful but likely unnecessary (4/6, 67%). As all chat interactions were in English, participants who were not confident in their English communication skills felt that their use of this feature would be limited (3/6, 50%). Others felt reminded of troubleshooting cable services rather than feeling connected to their provider. It was emphasized that any opportunity to improve connections to their providers through the app would be appreciated.

Discussion
Principal Findings and Implications
This study provides an applied example of a DHI for Chinese Canadian PCa survivors, which is based on broader principles of collectivism and relationality from Indigenous and Black feminist theory. Our initial aim was to co-design a cultural adaptation of the NED Clinic to provide compassionate care and meet the unmet needs of Chinese Canadian PCa survivors. Participants were also asked if they might find this information helpful. Although the majority (4/6, 67%) said no, those who said yes (2/6, 33%) were keen on having this information, especially if they needed to travel outside of Canada.

The questionnaire and assessment were generally deemed to be acceptable by most participants (4/6, 33%), with several notable dissents (2/6, 33%). The EPIC-CP question regarding hormonal function was highlighted as confusing by some because the connection between hormonal function and fatigue was not readily apparent. The spiritual domain in the needs assessment was flagged, as some thought that it would not be appropriately addressed by the nurse. Those who felt uncomfortable with this domain noted that they would prefer speaking about these needs to a spiritual leader. Agreement on appropriate response times also varied.

The chat function was deemed possibly helpful but likely unnecessary (4/6, 67%). As all chat interactions were in English, participants who were not confident in their English communication skills felt that their use of this feature would be limited (3/6, 50%). Others felt reminded of troubleshooting cable services rather than feeling connected to their provider. It was emphasized that any opportunity to improve connections to their providers through the app would be appreciated.

As resources could be accessed on demand, some indicated that more would be beneficial. However, other participants expressed that the number displayed in the prototype were more than sufficient, reflecting our previous study findings on the bifurcated information-seeking behaviors of Chinese Canadian PCa survivors. Participants were also asked if they might find having their imaging results helpful. Although the majority (4/6, 67%) said no, those who said yes (2/6, 33%) were keen on having this information, especially if they needed to travel outside of Canada.

The questionnaire and assessment were generally deemed to be acceptable by most participants (4/6, 33%), with several notable dissents (2/6, 33%). The EPIC-CP question regarding hormonal function was highlighted as confusing by some because the connection between hormonal function and fatigue was not readily apparent. The spiritual domain in the needs assessment was flagged, as some thought that it would not be appropriately addressed by the nurse. Those who felt uncomfortable with this domain noted that they would prefer speaking about these needs to a spiritual leader. Agreement on appropriate response times also varied.

The chat function was deemed possibly helpful but likely unnecessary (4/6, 67%). As all chat interactions were in English, participants who were not confident in their English communication skills felt that their use of this feature would be limited (3/6, 50%). Others felt reminded of troubleshooting cable services rather than feeling connected to their provider. It was emphasized that any opportunity to improve connections to their providers through the app would be appreciated.

Discussion
Principal Findings and Implications
This study provides an applied example of a DHI for Chinese Canadian PCa survivors, which is based on broader principles of collectivism and relationality from Indigenous and Black feminist theory. Our initial aim was to co-design a cultural adaptation of the NED Clinic to provide compassionate care and meet the unmet needs of Chinese Canadian PCa survivors via digital health.

However, attending to cultural adaptation theory and the lived realities of settler colonialism identified gaps to interweave Indigenous and Black feminist teachings. We began by synthesizing design principles that surfaced as critical to our participants and their feelings of comfort and safety when receiving follow-up care. This allowed us to leverage digital health to strengthen relations between the survivor and their providers; improve accessibility to resources; and honor desires for relationality, accountability, and care [46,47]. Rather than adapting by defining Chinese Canadian culture, we co-designed...
to intervene in structural causes of health inequities created by settler colonial culture instead [21,48].

We applied *Etuaptmun* by interweaving strengths from different ways of being and knowing, including those from Indigenous, Western, Chinese, and Black feminist traditions in relation to PCa follow-up and virtual care [27,30,31]. These included prioritizing relational care, accounting for the use of prostate-specific antigen screening as a recurrence monitoring tool, and the benefits of supportive care programs to create adaptation features [30,49]. The EPIC-CP validated questionnaire is a key part of clinical follow-up care, as it allows clinicians to identify possible areas of concern during follow-up [50]. The needs assessment addresses domains beyond clinical care, reflecting the holistic nature of the medicine wheel [51]. Access to resources includes education and guidance for the self-management of concerns across multiple domains. The app presents a “care contract” in the form of a schedule that clearly states the “terms” and dates of the user’s follow-ups [52]. It also respects the user’s privacy by providing access and allowing them to share their PHI on their terms [53]. Only key inputs are communicated for triage and response. Finally, language access is built into the app as a question of communication accessibility, rather than only culture.

This design approach and these features do not deny the fact that culture is a real influence and can be a source of strength in many peoples’ lives. However, we must go beyond implicating culture when designing DHIs for communities made vulnerable and instead address the overarching and underlying structures that create health inequities. Our design approach looked “up” at these structural causes rather than looking “down” and museumizing culture for participants through cultural sensitivity and competency. We demonstrate that a structural approach that applies teachings such as cultural safety and intersectionality can result in DHIs that are found to be acceptable, appropriate, and feasible for use while still leaving room for users to self-define and practice culture on their own terms. We are supporting, not replacing, the *labor* and *acts* of caring with digital health. Beginning with a paradigm shift opened a window to design for collective care, a scalable opportunity to benefit communities beyond Chinese Canadians with this *Ned Nurse* patient-facing app adaptation.

**Strengths and Limitations**

We have created the first “cultural” adaptation of a PCa follow-up care application for Chinese Canadian survivors. We extended the accessibility of this prototype by offering it in 3 language versions and tested its validity through member checking by returning it to participants who had provided their experiences and expertise as part of the first phase of this project. The findings should be considered with some limitations. Our sample does not fully represent the Chinese Canadian PCa community, as the heterogeneity of the community makes it difficult to recruit a fully representative sample [42]. User testing did not differentiate between results derived from users who interacted with the app themselves and users who directed a facilitator to perform actions on their behalf. However, all participants received the same set of instructions to apply the think-aloud method. A broad description of our theoretical stance, setting and place, methods, and results are provided to enhance understanding. We think of and encourage the transferability of this research as to how it might be made meaningful (ie, valid) for other communities in places where they may be subject to similar constructs and patterns of oppression [32]. Finally, this study does not include the provider perspective, although *Ned* was developed with clinicians who provide follow-up care for patients from this community. Future studies should examine the clinician’s perspective on the design and development of similar DHIs, including provision of care through these apps, acceptability and feasibility, and implementation readiness.

**Conclusions**

This study demonstrates the relationality of Indigenous and Black feminist ontologies, epistemologies, and methodologies to digital health design by providing a worked example of its empirical use for an adaptation of a PCa follow-up care app, the *Ned Nurse* Clinic, for Chinese Canadian PCa survivors. We applied UCD principles to develop a prototype design that supports the relational act of caring through digital technology by identifying structures that create inequities in the experiences of this community of survivors and designing to intervene and provide accessible, connected care instead. We hope that this prototype serves as a tool to help regenerate places of caring, as we have learned from Indigenous and Black feminist scholars’ teachings on power, place, and digital technologies.

**Acknowledgments**

We are deeply grateful for the time and effort gifted to us by the community members who shared their stories. We acknowledge and appreciate the staff and members of the Prostate Cancer Supportive Care Program, including Dr C Higano and Ms M Sundar, and the Richmond Chinese Prostate Cancer Support Group (列治文華人前列腺癌支援網絡), led by Mr D Leung and advised by Dr W Yu Ko. We thank the *Ned* team at the Centre for Digital Therapeutics (University Health Network) for their efforts and commitment to delivering a DHI for Canadian prostate cancer survivors. We give special thanks to Laura Parente for her incisive and human-centered design guidance. The study was funded by the Canadian Institutes of Health Research (CIHR) through the CIHR Canadian Cancer Society Survivorship Grant (CCS: 706713; CIHR: 168606) and the Canada Graduate Scholarship program, as well as by AGE-WELL Networks of Centres of Excellence through the Early Professionals & Inspired Careers in AgeTech (EPIC-AT) Fellowship and Graduate Award. These agencies were not involved in the design or analysis of this study.
Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions
KY, QP, ATB, ML, and WYK contributed to project conceptualization and study design. KY, RL, and TX contributed to data collection and analysis. KY prepared the first manuscript draft, with contributions from TX, WYK, ATB, ML, and QP. All authors contributed, reviewed, and approved the manuscript.

Conflicts of Interest
QP and the University Health Network (Toronto, Ontario) jointly own intellectual property rights to the Ned app. Under the respective agreements with their organizations, QP is entitled to personally benefit from any commercial use of the intellectual property.

Multimedia Appendix 1
Semistructured usability testing interview guide.

References


cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Original Paper

Effectiveness and User Perception of an In-Vehicle Voice Warning for Hypoglycemia: Development and Feasibility Trial

Caterina Bérubé1, PhD; Vera Franziska Lehmann1,2, PhD; Martin Maritsch1, PhD; Mathias Kraus3, Prof Dr; Stefan Feuerriegel4, Prof Dr; Felix Wortmann1,5, Prof Dr; Thomas Züger1,2,6, MD; Christoph Stettler1,2, Prof Dr; Elgar Fleisch1,5, Prof Dr; A Baki Kocaballi7, PhD; Tobias Kowatsch1,5,8,9, Prof Dr

1Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
2Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Bern University Hospital, University of Bern, Bern, Switzerland
3School of Business, Economics and Society, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberg, Germany
4School of Management, Ludwig-Maximilians-Universität München, Munich, Germany
5Centre for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland
6Department of Endocrinology and Metabolic Diseases, Kantonsspital Olten, Olten, Switzerland
7School of Computer Science, University of Technology Sydney, Sydney, Australia
8Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
9School of Medicine, University of St Gallen, St Gallen, Switzerland

Corresponding Author:
Tobias Kowatsch, Prof Dr
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
WEV G 216
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 44 632 94 88
Email: tkowatsch@ethz.ch

Abstract

Background: Hypoglycemia is a frequent and acute complication in type 1 diabetes mellitus (T1DM) and is associated with a higher risk of car mishaps. Currently, hypoglycemia can be detected and signaled through flash glucose monitoring or continuous glucose monitoring devices, which require manual and visual interaction, thereby removing the focus of attention from the driving task. Hypoglycemia causes a decrease in attention, thereby challenging the safety of using such devices behind the wheel. Here, we present an investigation of a hands-free technology—a voice warning that can potentially be delivered via an in-vehicle voice assistant.

Objective: This study aims to investigate the feasibility of an in-vehicle voice warning for hypoglycemia, evaluating both its effectiveness and user perception.

Methods: We designed a voice warning and evaluated it in 3 studies. In all studies, participants received a voice warning while driving. Study 0 (n=10) assessed the feasibility of using a voice warning with healthy participants driving in a simulator. Study 1 (n=18) assessed the voice warning in participants with T1DM. Study 2 (n=20) assessed the voice warning in participants with T1DM undergoing hypoglycemia while driving in a real car. We measured participants’ self-reported perception of the voice warning (with a user experience scale in study 0 and with acceptance, alliance, and trust scales in studies 1 and 2) and compliance behavior (whether they stopped the car and reaction time). In addition, we assessed technology affinity and collected the participants’ verbal feedback.

Results: Technology affinity was similar across studies and approximately 70% of the maximal value. Perception measure of the voice warning was approximately 62% to 78% in the simulated driving and 34% to 56% in real-world driving. Perception correlated with technology affinity on specific constructs (e.g., Affinity for Technology Interaction score and intention to use, optimism and performance expectancy, behavioral intention, Session Alliance Inventory score, innovativeness and hedonic motivation, and negative correlations between discomfort and behavioral intention and discomfort and competence trust; all
P<.05). Compliance was 100% in all studies, whereas reaction time was higher in study 1 (mean 23, SD 5.2 seconds) than in study 0 (mean 12.6, SD 5.7 seconds) and study 2 (mean 14.6, SD 4.3 seconds). Finally, verbal feedback showed that the participants preferred the voice warning to be less verbose and interactive.

Conclusions: This is the first study to investigate the feasibility of an in-vehicle voice warning for hypoglycemia. Drivers find such an implementation useful and effective in a simulated environment, but improvements are needed in the real-world driving context. This study is a kickoff for the use of in-vehicle voice assistants for digital health interventions.

(JMIR Hum Factors 2024;11:e42823) doi:10.2196/42823

KEYWORDS
hypoglycemia; type-1 diabetes mellitus; in-vehicle voice assistant; voice interface; voice warning; digital health intervention; mobile phone

Introduction

Background

Type 1 diabetes mellitus (T1DM) is a chronic condition caused by an inability of the pancreas to produce insulin and requires lifelong insulin therapy [1]. Hypoglycemia, also known as low blood glucose, is a frequent and acute complication in patients with T1DM [2,3]. Symptoms range from autonomic reactions such as trembling, anxiety, and hunger (ie, mild hypoglycemia) to neuroglycopenic reactions such as vision impairment, weakness, or cognitive impairments (ie, severe hypoglycemia) [2,4-6]. Hypoglycemia is a major issue in the context of driving: research has shown that hypoglycemia is associated with a higher risk of car mishaps [7-9]. In fact, drivers experiencing hypoglycemia are recommended by the local authorities [10] to stop the car and treat their condition. However, drivers do not always comply with these recommendations [11,12]. Thus, to help reduce hypoglycemia-related car accidents, there should be an effective warning that informs the driver about an upcoming hypoglycemic episode and supports the driver in coping with the situation. Currently, hypoglycemia can be detected and signaled through flash glucose monitoring (FGM) or continuous glucose monitoring (CGM) devices (ie, wearable receivers connected to a sensor inserted in the subcutaneous tissue of the arm or abdomen) [13]. These allow for glucose monitoring by displaying the values either continuously (ie, CGM) or upon active retrieval (ie, FGM) and deliver alerts in the form of a tone or vibration in case of out-of-range values. However, these devices present limitations in the context of driving. For instance, FGM needs to be held close to the sensor to transfer the data from the subcutaneous sensors to the monitoring device, that is, the driver needs to actively engage in a manual gesture to access the glucose value and to look at a visual display moving the focus of attention from the driving task. In contrast, allowing the drivers to receive an alert in a hands-free mode will facilitate warning reception [14] and lower worry associated with driving with T1DM [15]. However, hypoglycemia is known to cause a decrease in attention [2,4-6], thereby challenging the effectiveness of such devices. As 90% of road accidents are caused by human error, the European Commission has set new safety technologies as mandatory equipment for vehicles as of 2022 (eg, driver drowsiness and distraction warnings and speed assistance) [16]. In-vehicle warning systems for impaired driver states, such as fatigue [17], distraction [18], and breath alcohol concentration [19], are increasingly being developed. However, to the best of our knowledge, there is no existing implementation for hypoglycemia. Such technology would be aligned with the “healing car” concept [20], where vehicles become environments promoting well-being for passengers, including ergonomic seats, ambient lighting, relaxation exercises [21], and detection of health-critical states [22]. This concept is still in its early stages, but it may become a standard in car manufacturing in the future. So far, the only attempts of in-vehicle glucose monitoring are either only proof of concept without user validation [23] or conceptual work [24]. However, the online community clearly expressed a need for in-vehicle glucose monitoring and warning [25].

A growing number of automotive companies are introducing voice assistance technology into their products [26,27]. Voice assistants add value not only for the associated consumer experience but also for their greater safety. Indeed, vocal interactions have been observed to be the least cognitively demanding while driving compared with visual and haptic interactions [28,29]. Moreover, voice assistants are increasingly being implemented to deliver digital health interventions [30-33]. Although research is still in its infancy, efforts have been made to develop voice-based conversational agents to monitor and support individuals with chronic diseases such as cancer, cardiovascular diseases, cognitive disorders, or diabetes [30]. Other recent examples include prevention of excessive alcohol consumption [34], health education and monitoring, physical and mental exercise, and nutrition [35]. Furthermore, a voice assistant delivering a warning is a form of proactive behavior initiated by the computer rather than the user [36,37]. In-vehicle voice assistants can provide personalized and adaptive suggestions, but users may ignore proactive behavior if it is inopportune, violates privacy, or distracts from driving [38-40]. However, emergencies are the most suitable context for proactive behavior that violates privacy [39].

Objectives

Therefore, we investigate the feasibility of an in-vehicle voice warning delivered by a built-in voice assistant to alert and support drivers with T1DM during hypoglycemia. To the best of our knowledge, there have been no investigations on safe and effective in-vehicle hypoglycemia warnings to support drivers with T1DM or on the perception of such technology. Thus, we sought to answer the following research questions (RQs):

https://humanfactors.jmir.org/2024/1/e42823
• RQ1: How do drivers perceive an in-vehicle voice warning for hypoglycemia while driving?
• RQ2: How effective is an in-vehicle voice warning in prompting drivers to cope with hypoglycemia?

RQ1 refers to the attitude of drivers toward the warning, whereas RQ2 refers to the driver’s compliance behavior once the warning is delivered. Answering these RQs will allow us to conclude on the feasibility of an in-vehicle voice warning for hypoglycemia. To control for individual factors influencing the perception of the warning [41], we also assessed technology affinity.

Methods

Study 0: Preliminary Assessment With Healthy Individuals in Simulated Driving

Driving Setting
Participants performed the task in a driving simulator (Carnetsoft Inc) with 3 monitors displaying the front, left, and right views. The central monitor also showed the cockpit and navigation arrows. The participants used a steering wheel and pedals (Logitech Driving Force G29) to control the simulator, which was set to automatic (ie, no clutch or gear shifter). The simulator’s computer was connected to a stereo speaker with a subwoofer, which was kept at a constant volume. To control for driving difficulty, 3 environments were used: highway, countryside, and town, with the first and last being the least and most difficult, respectively.

In-Vehicle Voice Warning Simulation
Before testing a hypoglycemia voice warning with people with T1DM, we tested the concept of a car voice assistant as an interface between a dedicated monitoring system and the user with healthy participants. As the participants were not affected by hypoglycemia, the first version of the warning was a simulated low fuel warning (“The car needs a refill. Please pull over and turn off the engine”). Although not health related, it signaled an event of reasonable urgency that required safely stopping the car. Note that the participants were informed that this message aimed to ask them to stop the car as soon as possible and that they did not need to look for a gas station.

The warning was simulated using the Wizard-of-Oz method, where the conversational turns produced by the voice assistant were played by the experimenter [42] from a laptop using predefined keyboard keys. The turns were based on the Google Cloud text-to-speech engine, with a de-DE-Wavenet-C voice, a speed of 1.11 times the normal native speed of the specific voice, and a pitch of −1.20 semitones from the original pitch. The experimenter’s computer was connected to the same sound system as the driving simulators so that the voice warning could be heard as part of the driving simulation. No visuals were included.

Voice Warning Evaluation Measures
To assess the RQs, we assessed participants’ perception of the warning (self-reported through the modular evaluation of key Components of User Experience [meCUE]; 10 constructs evaluated on a 7-point Likert scale and a general evaluation evaluated on a 10-point scale [43,44]) and participant compliance with the warning (measured by the experimenter manually assessing if the participant would pull over and stop the car following the warning, and reaction time in seconds from the timestamp of the warning to the timestamp of the car fully stopped). As the perception of technology can be influenced by technology-related personality [45], we also measured technology affinity (measured by the Affinity for Technology Interaction [ATI], a 6-point Likert scale [46]). Finally, qualitative feedback was collected informally.

Evaluation Procedure
The participants were welcomed, informed about the procedure, and invited to sit in the simulator. The voice assistant introduced itself and invited the participants to familiarize themselves with the setting, including the 3 environments. The training also screened for motion sickness.

In the experimental session, participants drove 12 times, with 4 blocks of 3 drives each, for approximately 5 minutes per drive. The driving environment’s order and starting point varied to minimize habituation. The drive began when the voice assistant prompted participants to start the engine. A timer started to deliver the low fuel warning at either 100 or 200 seconds to add variation and minimize habituation effects. At the end of the session, participants completed the meCUE.

Data Analysis
Participants were characterized by sex, age, and driver’s license duration. The ATI was aggregated as a whole, and meCUE items were aggregated per construct. All reports were aggregated across the sample, with mean and SD. Compliance was coded as binary (0=not compliant, 1=compliant) and reported in terms of frequency. Reaction time was aggregated in seconds across participants and phases, with mean and SD.

Study 1: Assessment With Individuals With T1DM in Simulated Driving
Following the iterative approach described earlier, we conducted 3 exploratory iterations. This study was part of a clinical trial registered at ClinicalTrials.gov (NCT04035993).

Driving Setting
The driving setting was the same as in study 0.

In-Vehicle Voice Warning Simulation
On the basis of the results of study 0, we adapted the warning to hypoglycemia instead of low fuel, using the fewest conversational turns possible [47]. To ensure that the drivers were available, the voice assistant started with a receptivity check: “May I disturb you?”

We designed the warning based on the guidelines of the Swiss Diabetes Association [10], which recommends taking carbohydrates and stopping the car as soon as signs of hypoglycemia are noticed. To give the driver a sense of autonomy [48], we designed the warning to suggest eating carbohydrates rather than directly engaging in stopping the car. However, if the driver did not have carbohydrates, they were asked to pull over. On the basis of the feedback, we enhanced
the voice warning used in the following study to recommend pulling over directly (detailed conversation flow is available in Multimedia Appendix 1).

As in study 0, the warning was simulated with a Wizard-of-Oz method [42], and the turns were generated by recording the same voice. However, to reduce fatigue and cognitive load, we decreased the speed and pitch to 0.93 times the normal speed and −4.8 semitones from the original pitch, respectively. As in study 0, the experimenter would play the turns from a Microsoft Windows laptop using predefined keyboard keys to play prerecorded voice sounds. However, in study 1, the laptop program included a visualization mirrored on a smartphone. The visuals consisted of a blue circle that gradually faded in and out when the voice assistant was speaking. As in study 0, the experimenter’s computer was connected to the same sound system as the driving simulators, so that the voice assistant could be heard as part of the driving simulation.

**Voice Warning Evaluation Measures**

Perception assessment focused on evaluating the voice assistant as a trustworthy driving companion. Specifically, participants completed the Acceptance and Use of Technology (AUT) questionnaire [49,50], the Session Alliance Inventory (SAI) [51], and the Emotional Trust and Competence Trust subscales (henceforth Trust) of the Trust and Adoption questionnaire [52].

To assess technology affinity, participants completed the streamlined scale of the Technology Readiness Index (TRI 2.0) [53]. Items were rated on a 5-point Likert scale (ie, 1=totally disagree, 5=totally agree). We also added a question on whether the participants had previous experience with in-vehicle voice assistants (ie, “Have you already had experience with in-vehicle voice assistants?” with a yes or no answer).

Finally, to obtain qualitative and more in-depth feedback for improvement, we conducted a semistructured interview about their experience with the warning (the interview questions are provided in Multimedia Appendix 2).

**Evaluation Procedure**

The procedure was the same as in study 0, except that participants drove only once for 5 minutes (the evaluation procedure is detailed in Multimedia Appendix 3). Before driving, we ensured that the participants had normal blood glucose levels (5-8 mmol/L).

**Data Analysis**

The sample of participants was characterized by sex, mean age, and mean duration of their driver’s license before the study. TRI and SAI were aggregated as a whole, and the AUT and Trust items were aggregated per construct. Scores from the negatively formulated questionnaire items were inverted. Previous experience with an in-vehicle voice assistant was reported in terms of frequency. All these reports were aggregated across participants of each iteration, with mean and SD.

To further explain results in perception, they were associated with technology affinity measures. The difference in perception between participants with and without experience with an in-vehicle voice assistant was tested using a 2-sided t test, and it was correlated with the TRI constructs using a Pearson test.

Compliance was defined as whether the participant would comply with the warning and was coded as binary (0=did not comply, 1=complied). Reaction time was aggregated in seconds with mean and SD. Compliance behavior was aggregated across participants of each iteration.

Feedback was summarized in positive and negative topics, with a focus on the most prominent suggestions for improvement. Feedback was aggregated across participants of each iteration.

**Study 2: Assessment With Individuals With T1DM in Real-World Driving Undergoing Hypoglycemia**

Following the iterative approach described earlier, we conducted 2 exploratory iterations. This study was part of a clinical trial registered at ClinicalTrials.gov (NCT04569630).

**Driving Setting**

Participants drove in Volkswagen Touran on a closed circuit accompanied by a driving instructor. Dual pedals allowed the driving instructor to intervene and stop the car if necessary. The driving environments on the test track corresponded to the environments of the driving simulator used in the previous studies. Straight paths, turns, crossroads, stop signs, and a pedestrian crossing with a doll were used to implement the highway, countryside, and town scenarios. Artificial obstacles (eg, boxes and lines of traffic pylons) were used to simulate the traffic.

**In-Vehicle Voice Warning Simulation**

On the basis of the participant feedback from study 1, we revised the voice warning and addressed low trust ratings by explaining the cause of the warning. We simulated driving behavior as a trigger to detect hypoglycemia while driving, as in the study by Lehmann et al [54]. We created 2 variations of the simplified hypoglycemia notification—one with a statement of the cause (driving behavior) and one without. The final recommendation was reformulated as stricter but less directive than that in study 1.

In the second iteration, we simplified the conversational flow by removing the receptivity check (“May I disturb you?”) and the final recommendation (Multimedia Appendix 1 provides the conversation flow).

We used the Wizard-of-Oz method to simulate the warning, as in studies 0 and 1. We implemented the voice assistant in a smartphone with the same voice as in study 1. However, the experimenter had to control it remotely (outside the car), so we implemented the interaction in a smartphone app controlled by a remote desktop application. The experimenter used the smartphone screen to control the voice warning delivery; therefore, no visualization was included. Because of network-related slowdowns in the remote control, we used a combination of remote control and speech-to-text programing.

**Voice Warning Evaluation Measures**

All measures were the same as in study 1. Reaction time was calculated from the warning onset until the car reached a velocity of 0. In addition, at the end of the experiment, we
included a questionnaire item asking which of the 2 types of warning they preferred, that is, the warning including a statement of the cause that triggered the warning or the one without it, or if they would not use either of them.

**Evaluation Procedure**

After welcoming participants and explaining the procedure and simulated voice assistant, the voice assistant introduced itself as an in-vehicle assistant to support drivers with hypoglycemia. The participants then completed a training drive.

The warning was delivered at different stages of hypoglycemia (see the study by Lehmann et al [54]). Drive blocks were defined based on blood glucose levels. In the first phase, the participants drove at normal glucose (5-8 mmol/L). In the second phase, blood glucose level was progressively lowered below the moderate hypoglycemia threshold (3.0 mmol/L) to a target range of 2.0 to 2.5 mmol/L. In the third phase, moderate hypoglycemia was maintained. In the fourth phase, participants drove again with normal blood glucose levels (Multimedia Appendix 3).

To explore the effect of blood glucose level on warning perception and compliance, we delivered a warning at the end of the last drive of each phase. Participants received 2 warnings with an explanation and 2 without, in randomized order.

**Data Analysis**

Data analysis was carried out as in study 1.

**Ethical Considerations**

Study 0 was approved by the Ethics Board of ETH Zürich, Switzerland (2019-N-32), and study 1 and study 2 were approved within the context of the HEADWIND study by the cantonal ethics commission of Bern, Switzerland (2020-00685 and 2021-02381, respectively). Study 1 and study 2 are available at ClinicalTrials.gov (NCT04035993 and NCT04569630, respectively). All participants provided written informed consent.

**Results**

**Study 0: Preliminary Assessment With Healthy Individuals in Simulated Driving**

Results are summarized in Figure 1.

**Recruitment and Participants**

We recruited 11 healthy individuals with a valid driver’s license via a web advertisement (ie, University of Zurich marketplace). One participant was excluded owing to simulator sickness. Thus, we included 10 participants (n=4, 40% female; n=6, 60% male) with an average age of 30.4 (SD 7.8; range 23-47) years and holding a license for 11 (SD 7.5; range 2-26) years, on average.

**Technology Affinity Measure**

Participants showed a mean ATI of 4.2 (SD 1; Cronbach α=.91), which is 70% of the maximal value.

**Perception Measure**

The meCUE (Cronbach α=.7) revealed a mean overall evaluation of 6.4 (SD 1.6), which is 64% of the maximal value. Moreover, the highest mean values were achieved for usability (mean 6.2, SD 0.6, 89%) and usefulness (mean 5.6, SD 0.9, 80%), whereas lower values were observed for commitment (mean 1.5, SD 0.4, 21%), positive emotions (mean 2.7, SD 1.1, 39%), negative emotions (mean 2.7, SD 1, 39%), intention to use (mean 3.1, SD 1.1, 44%), and product loyalty (mean 2.6, SD 0.7, 37%). A low value for negative emotions reflects a more positive evaluation.

To explain the perception results with the technology affinity measure, we correlated each meCUE construct with ATI. We observed a correlation between ATI and intention to use (p=0.70; P=.02). All the other correlations were not significant at the .05 level.

**Compliance Measure**

All the participants complied with the warning and stopped the car. Participants took 12.6 (SD 5.7) seconds on average.
**Qualitative Feedback**

Finally, some participants reported that the voice assistant spoke too fast to deliver information during a driving task without being distracting.

**Study 1: Assessment With Individuals With T1DM in Simulated Driving**

Results are summarized in Figures 2 and 3.

**Figure 2.** Violin plots of (A) count of previous experience, (B) score values across the constructs of Technology Readiness Index (TRI; min=1, max=5), (C) score values across the constructs of Acceptance and Use of Technology (AUT; min=1, max=5), (D) Session Alliance Inventory (SAI) scores (min=1, max=5), (E) Trust scores (min=1, max=5), and (F) reaction time across iterations in study 1 (n=18). The dots represent the group means; the dashed line represents the overall mean within an iteration. RT: reaction time; sec: seconds.

**Figure 3.** Thematic summary of participants' feedback in study 1 (n=18).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Content to add</th>
<th>Content to remove</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>First (n=9)</td>
<td>Always suggest to stop the car (n=4)</td>
<td>Asking to wait 30 minutes may be irrelevant (n=2)</td>
<td>Not dropping but too low (n=3)</td>
</tr>
<tr>
<td>Second (n=7)</td>
<td>Always suggest to check blood glucose (n=4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third (n=2)</td>
<td></td>
<td>Shorter warning (n=1)</td>
<td></td>
</tr>
</tbody>
</table>
**Recruitment and Participants**

We recruited 20 patients with T1DM from the Department of Diabetes, Endocrinology, Nutritional Medicine, and Metabolism at the Bern University Hospital. Participants needed functional insulin treatment, good insulin self-management knowledge, a driver’s license, and active driving in the past 6 months. We excluded one participant owing to simulator sickness and one participant owing to technical errors in the warning delivery. This resulted in a total of 18 participants (n=6, 33% female and n=12, 67% male; mean age 31.4, SD 7, range 24-44 years; mean driver’s license age 13, SD 7.5, range 4.5-28.6 years). The first iteration had 9 participants, the second iteration had 7, and the third iteration had 2 participants. Although the last iteration’s sample size was small, it provided useful feedback to improve the warning for study 2.

**Technology Affinity Measure**

Seven participants had previous experience with an in-vehicle voice assistant (n=2, 28% in the first iteration; n=3, 43% in the second iteration, and n=2, 28% in the third iteration). TRI was 3.4 (SD 0.6, Cronbach $\alpha=0.85$), or 68% of the maximum. Specifically, TRI was 3.5 in the first and second iterations (SD 0.6 and 0.7, respectively) and 2.7 in the third iteration (SD 0.9).

**Perception Measure**

Perception averaged 3.9 out of 5 (SD 0.8, 78%) and remained stable across iterations. Average AUT (Cronbach $\alpha=0.81$) values were 3.8 (SD 1) in the first iteration, 4 (SD 0.7) in the second iteration, and 3.8 (SD 0.8) in the third iteration. Effort expectancy and facilitating conditions had the highest values across all iterations, whereas behavioral intention always had the lowest values.

SAI (Cronbach $\alpha=0.79$) averaged 3.7 out of 5 (SD 0.5, 74%) and increased slightly over the iterations, from 3.4 (SD 0.7) in the first iteration to 3.6 (SD 0.6) in the second iteration to 4.1 (SD 0.1) in the third iteration.

Trust (Cronbach $\alpha=0.8$) averaged 3.1 out of 5 (SD 0.7, 62%), was stable across constructs, and had the lowest values of the 3 perception measures. Trust averaged 3.1 (SD 0.8) in the first iteration, 2.9 (SD 0.6) in the second iteration, and 3.3 (SD 0.6) in the third iteration.

To explain the perception results with technology affinity, we tested the difference in perception (AUT, SAI, and Trust) between participants with and without previous experience with in-vehicle voice assistants. The means of all constructs, excluding facilitating conditions, were slightly higher for participants with previous experience. However, a 2-sided $t$ test revealed no significant result (ie, $P>.05$).

We also correlated perception with TRI and observed a correlation between the optimism construct and performance expectancy ($P=0.49$; $P=.04$), behavioral intention ($P=0.52$; $P=.03$), and SAI ($P=0.57$; $P=.01$). All the other correlations were not significant ($P>.05$).

**Compliance Measure**

All the participants complied with the warning. In the first iteration, all participants answered *yes* or *no* to the receptivity check (“May I disturb you?”) and when asked if they had carbohydrates on hand. Five of the 9 participants answered *yes* to the latter question, although they did not. Two of those 9 participants stopped the car although they were not explicitly advised to do so. In the second iteration, all participants answered the prompts with *yes* and stopped the car as advised. One participant gave an affirmative *mhm* when asked, “May I disturb you?” during the hypoglycemic phase but were otherwise compliant. Because we used the Wizard-of-Oz method, the experimenter interpreted the affirmation. However, a current voice assistant might have interpreted it as an error. In the third iteration, both participants answered the prompts with *yes* and stopped the car. Across iterations, compliance took approximately 22 seconds. In particular, compliance took approximately 20 (mean 20.7, SD 6.2) seconds in the first iteration, approximately 17 (mean 16.7, SD 1.2) seconds in the second iteration, and approximately 31 (mean 31.7, SD 10.6) seconds in the third iteration.

**Qualitative Feedback**

Participants judged the voice warning as pleasant, simple, and as clear and efficient (n=15, n=11, and n=13, respectively). The topics for improvement are summarized in Figure 3. Note that these results are best understood when compared with Multimedia Appendix 1.

Given that Trust showed the lowest values in the first iteration, in comparison with the other perception measures, we decided to specifically ask participants, in our second and third iterations, what would help them trust the warning more. Of the 9 participants included in both the second and third iterations, 5 (55%) said they would just need to have a prolonged experience with the warning, whereas 3 (33%) said they would need to know what kind of data is used to infer that the driver is about to experience hypoglycemia. One participant did not know what would improve their trust.

**Study 2: Assessment With Individuals With T1DM in Real-World Driving Undergoing Hypoglycemia**

Results are summarized in Figures 4 and 5.
Figure 4. Violin plots of (A) count of previous experience, (B) score values across the constructs of Technology Readiness Index (TRI; min=1, max=5), (C) score values across the constructs of Acceptance and Use of Technology (AUT; min=1, max=5), (D) Session Alliance Inventory (SAI) scores (min=1, max=5), (E) Trust scores (min=1, max=5), and (F) reaction time across iterations in study 2 (n=20). The dots represent the group means; the dashed line represents the overall mean within an iteration. RT: reaction time; sec: seconds.

Figure 5. Thematic summary of participants’ feedback in study 2 (n=20).
Recruitment and Participants
The recruitment procedure was the same as in study 2. We recruited 21 individuals, and 1 participant was excluded owing to data loss. Thus, we included 20 participants (n=3, 15% female and n=17, 85% male; mean age 40.9, SD 10.6, range 23-57 years; and holding a license on average since 23.7, SD 11.1, range 3.1-42.4 years). The first iteration included a sample of 9 participants and the second iteration included a sample of 11 participants.

Technology Affinity Measure
The pretest measurements revealed that 25% (5/20) of the participants had previous experience with an in-vehicle voice assistant (2 in the first iteration, and 3 in the second iteration), whereas TRI was on average 3.4 (SD 0.7; Cronbach α=.44), which is 68% of the maximal value. In particular, TRI was 3.4 in the first iteration (SD 0.8), and 3.3 in the second iteration (SD 0.7).

Perception Measure
The overall perception score was 1.7 out of 5 (SD 1.3, 34%). The results showed a slight increase in mean AUT (Cronbach α=.95) and Trust (Cronbach α=.85) values between the first and the second iteration, whereas SAI (Cronbach α=.80) showed a slight decrease. AUT also showed a considerable increase in SD. In particular, AUT values were on average 1.4 (SD 1) in the first iteration, and 1.9 (SD 1.6) in the second iteration; SAI was overall 2.8 out of 5 (SD 0.8, 56%). Values were on average 3.1 (SD 0.7) in the first iteration and 2.5 (SD 0.9) in the second iteration; Trust values were on average 1.4 (SD 0.8) in the first iteration, and 2.3 (SD 1.1) in the second iteration.

Similar to study 1, to explain the perception results with the technology affinity measure, we tested the difference in perception (ie, AUT, SAI, and Trust) among participants who had previous experience with in-vehicle voice assistants and those who did not. The means of all perception measures, excluding SAI, were consistently slightly higher in the second iteration. The means of all constructs, excluding SAI, were consistently slightly higher for participants who had previous experience with in-vehicle voice assistants. However, a 2-sided t test revealed no significant result (ie, P>.05). When correlating each perception measure with TRI, we observed a correlation between innovativeness and hedonic motivation (p=.052; P=.02), a negative correlation between discomfort and behavioral intention (p=.046; P=.04), and a negative correlation between discomfort and competence trust (p=.045; P=.05). All the other correlations were not significant (P>.05).

Compliance Measure
All the participants complied with the warning. Two drivers were excluded: one participant stopped once before the warning was delivered and data from one drive of one participant was lost. The results showed that the reaction time does not seem to vary across glycemic phases and, although minimal, there is a tendency for the reaction time to increase in the second iteration. Participants took 13.6 (SD 4.5) seconds in the first iteration and 15.5 (SD 4.1) seconds in the second iteration.

Preference for the Disclosure of the Triggering Cause
One participant was excluded because of data loss. The results showed that although 10 participants preferred when the warning was delivered with an explanation for the warning being triggered (in this case, driving behavior), 8 participants preferred it without the explanation. One participant stated that they would not use this in-vehicle voice warning either way.

Qualitative Feedback
In general, and similar to study 1, the participants found the communication style pleasant and efficient (n=4 and n=5, respectively). The topics for improvement are summarized in Figure 5. Note that these results are best understood when compared with Multimedia Appendix 1.

Discussion
Principal Findings
Most participants had not previously used an in-vehicle voice assistant, and technology affinity was similar across studies. In general, the voice warning elicited a positive perception, although the perception values were lower in the real-car study. In addition, participants complied with the warning in all studies, and reaction times were shorter in the real-car study than in the simulator study. Finally, the participants preferred the voice warning to be less verbose and prompt fewer interactions with the driver.

Technology Affinity
Although we did not observe a significant effect on the perception of the warning, we suspect that the participants may have experienced a double novelty: using a voice assistant while driving and experiencing a warning from an in-vehicle voice assistant. Thus, future research should include a more balanced sample and compare the perception of a voice assistant–based warning with a standard warning (eg, an acoustic tone). Moreover, although we cannot directly compare ATI (used in study 0) with TRI (used in study 1 and study 2), we can observe that technology affinity was similar across studies. Although ATI showed a mean of 4.2 over 7 (60%), TRI showed a mean of 3.4 over 5 both in study 1 and study 2 (68%). The change in technology affinity measure was the result of an internal discussion between the coauthors, and we recommend the scientific community to use TRI in future research, as it is more widely used and focuses not only on the interaction but also on the general attitudes toward new technologies.

Perception
We observed that AUT, SAI, and Trust values were higher in study 1 (simulated driving) than in study 2 (real-world driving). This evaluation might have been influenced by the driving setting. There can be 2 possible reasons. First, participants may have found the warning to be more distracting in the real car than in the simulator. However, research shows that drivers are more in control in real-world driving than in simulated driving [55]. Second, the technical difficulties in controlling the driver-assistant interaction owing to network slowdowns might have affected the user experience, and thus the perception measures. Future Wizard-of-Oz studies may account for this.
methodological weakness with a more accurate text-to-speech technology, avoiding remote control, and reducing interactions. In addition, TRI seemed to have influenced behavioral intention (AUT) but did not consistently influence the other perception measures (ie, other constructs of AUT, SAI, and Trust). Thus, participants may have been excited about the potential of the voice warning, but they may not have been happy with the actual experience of using it.

**Compliance**

The reaction times were short enough to ensure a timely reaction to the critical event. Blood glucose can change with a maximum rate of 0.22 mmol/L/min [56]. This means that someone driving with a normal glucose of 5.5 mmol/L might reach hypoglycemia (ie, 3.9 mmol/L) within a minimum of 7.5 minutes. Thus, although experiencing hypoglycemia while driving does not require an abrupt stop but rather a careful pullover maneuver and treating the condition, measuring reaction time provided an insight into the time required to take the first measure (ie, pullover). Interestingly, the reaction time was shorter in the real car (study 2) than in the simulator (study 0 and study 1). This difference may be attributed to the lack of traffic in study 2, which allowed the driver to pull over faster.

**Feedback**

Although we aimed to keep the warning conversational, participants preferred a more direct notification of the problem without specific recommendations (eg, recommending waiting until the blood glucose is at its normal level) or polite formulations (eg, asking for permission to talk). To the best of the author’s knowledge, there was no in-vehicle voice warning at the time of the study, and we mostly relied on the guidelines of the Swiss Diabetes Association [10], while keeping the conversation as simple as possible. The participants’ feedback allowed us to improve the warning in this direction.

**Implications and Future Directions**

**Hypoglycemia Warnings**

Reportedly, no research has been conducted for in-vehicle applications providing a hypoglycemia warning. However, smartphone apps for hypoglycemic events tracking have been investigated [57]. Although most of the research on glucose monitoring solutions conducted so far focused on diary apps rather than warning delivery, a pilot study on a smartphone-based hypoglycemia warning showed an improved hypoglycemia awareness and a reduction in daytime hypoglycemia [58] (other research is still in the phase of validation [59]). Future research should investigate such outcomes with an in-vehicle extension of this type of application.

**In-Vehicle Warnings**

Although there seems to be no related work testing the voice assistant of a private vehicle to deliver hypoglycemia warnings, there is a need for “driver-friendly” in-vehicle glucose monitoring solutions, expressed by the online community [25]. In particular, drivers with T1DM have contributed to the Nightscout Foundation [60], a nonprofit organization founded in 2014 and supporting open source technology for T1DM management, with the development of a data-sharing app, able to connect a car to a CGM, and display the glucose trends while driving on the dashboard of the private vehicle [25]. Moreover, there has been conceptual work manifesting the need for collaboration between automotive and medical industries to improve the safety of drivers with T1DM [24]. However, this work has not been followed by any implementation. Furthermore, no testing with the actual users has been conducted. Our work provides preliminary evidence, both in a simulated and a real-world environment.

Needless to say, recognizing hypoglycemia is only one part of glucose monitoring while driving; general imbalance of blood glucose (including hyperglycemia) can be problematic for the driver, if not dangerous [61]. Our work can be extended to hyperglycemia and, therefore, support further the safety of drivers with T1DM.

Finally, using the in-vehicle voice assistant to deliver a warning is compatible with current technology: not only are cars increasingly equipped with voice assistants [26,27] but also the automotive industry is aware of the relevance of using the upcoming “in-car proactivity” [62].

**Warning Escalation**

Our results showed 100% compliance in all 3 studies. This can only mean that the warning was clear enough for the participant to understand that it was time to pull over. That is, as all studies were run in a controlled setting, where an experimental team was present, and the participant knew they would be recommended to pull over eventually, we can safely assume that the experiments experienced a participant bias [63]. Thus, we cannot conclude that the warning was compelling enough to motivate the participants to comply (see the Limitations section). Nevertheless, the warning should be designed to allow for escalation, whereas in case the driver does not pull over in due time (eg, 2-3 min [56]) or explicitly rejects the warning, delayed reprompts with an increasingly severe tone would be delivered by the voice assistant (eg, “You are at risk of hypoglycemia. Please stop the car safely and check your blood sugar, then risk of hypoglycemia. Pull over now”).

**Hypoglycemia Detection**

Finally, in this paper, we focus on the interface between the hypoglycemia detection system and the driver, with the aim of visually distracting them as little as possible. Although the detection side is beyond the scope of this study, the designed warning is intended to be produced by a voice assistant built into the vehicle. Therefore, how a vehicle monitors blood sugar depends on the technology of the car. For instance, the aforementioned open source app displaying the glucose levels on the dashboard of a private vehicle [25] could be enhanced to connect with the in-vehicle voice assistant and use a voice warning instead of a visual one. Furthermore, research has been conducted on how to detect hypoglycemia from the car’s data [54] and from consumer-available wearable devices [64], with the argument that CGM devices can impose a social and financial burden on the individual.
Limitations and Strengths
Despite our best efforts, this research has 3 main limitations.
First, the studies included a relatively small sample size. However, this study includes 3 feasibility studies (ie, a preliminary study with healthy individuals and 2 feasibility studies with individuals with T1DM), and the research presented in this paper is intended to be understood as an iterative development of a hypoglycemic warning. As such, this research aimed to pioneer the use of in-vehicle voice assistants for a driver health-related warning, rather than draw conclusions to be generalized to the population with T1DM. Thus, although we included a total sample size of 48 individuals, each feasibility study provides insight into the changes required by the users, and we provide the scientific community with an opening to the design of in-vehicle voice assistant–based health-related warning. Furthermore, previous studies on digital health systems used a similar sample size [65-67]. Thus, we believe that although the sample size does not allow drawing conclusions on the interaction of drivers with T1DM with in-vehicle hypoglycemia warnings, it still reports pioneer research.

Second, the studies were conducted over a short period. The participants had only a short-term experience with the warning. Perception and compliance may therefore be influenced by the novelty of such an experience, whereas perception may stabilize with repeated experience [68]. Future research should investigate the user experience of the warning in a longitudinal study. Third, these studies did not control for all potentially confounding variables related to real-world traffic and driver’s priorities. For instance, both simulator and real-car experiments involved disadvantages: while assessing the warning in a simulator allowed a controlled and safe experiment, such a setting remains artificial and lacks external validity. In contrast, while testing it in a real car increased the ecological validity of the human-machine interaction, it did not allow for as much traffic and speed variation as was possible in the simulator. Future research should investigate the effects of real-world traffic on the perception of the warning and compliance behavior. Moreover, receiving a warning in the presence of a team of experimenters may have influenced the participant’s verbal and behavioral responses; participants knew they would receive the warning sooner or later and had no reason not to follow it (eg, ignoring the warning because of being late for an appointment). In a real situation, drivers may not respond as expected or may even ignore the warning. Future research should test such a warning in a more ecological context, for instance, in a field study where the driver may not fall for a participant bias [69].

Finally, as we aimed to test a voice warning, our studies used a Wizard-of-Oz methodology to avoid problems related to natural language processing. Note that our studies were conducted in German-speaking Switzerland, where the German accents easily vary from region to region. As this aspect was beyond the scope of our research, we did not implement a working voice assistant or account for potential fallback intents triggered by the voice assistant’s failure to understand the user. Future research should push this research further and examine the potential danger of delayed treatment of hypoglycemia owing to the voice assistant’s natural language processing errors.

Conclusions
Although hypoglycemia increases the risk of car mishaps [7,8], current solutions (eg, CGM and FGM) require visual human-machine interaction, which is inappropriate for an in-vehicle context. As voice assistants are increasingly present in private vehicles [26,27] and the European Commission fosters safety technologies inside the car [16], we propose to warn the driver of their critical health state through a voice assistant–based health warning. This paper reports on an iterative development and assessment of a hypoglycemia warning. In particular, we conducted in 3 studies: a preliminary study using a simulator with healthy participants, a test with individuals with T1DM in a simulator, and a test with individuals with T1DM in a real car. This gradual increase in authenticity in the experimental design allowed us to increase the ecological validity of our results while keeping experimental control. To the best of our knowledge, this is the first attempt of such a comprehensive feasibility assessment of an in-vehicle voice warning for hypoglycemia. Our results suggest that a voice warning can be useful, but that proactive behavior in voice assistants is still emerging and unfamiliar. We hope that these preliminary findings will foster future research to further develop in-vehicle hypoglycemia warnings.

Acknowledgments
The authors would like to thank the Swiss Federal Department of Defense, Civil Protection and Sports for providing the closed circuit used in study 2.

This work was partly funded by the Swiss National Science Foundation, Sinergia Project (CRSII5 183569).

Data Availability
The data supporting the findings of this study are available from the last author, TK, upon reasonable request.

Authors’ Contributions
CS and EF were responsible for the oversight, leadership, and management of the research activity and for funding acquisition. CB, MM, VFL, MK, SF, TZ, FW, CS, EF, and TK were responsible for the methodology. CB and TK were responsible for the formulation of the research goals and aims, conceptualizing the voice warning. CB and MM were responsible for developing the voice warnings. CB, MM, FW, and TK were responsible for developing the driving scenarios. FW, TK, and TZ were responsible...
for providing the driving simulator, the real car, and the closed circuit. CB and MM were responsible for recruiting participants for study 0. VFL, TZ, and CS were responsible for recruiting participants for studies 2 and 3. CB, MM, and VFL were responsible for data collection. CB was responsible for the data analysis and presentation and the first draft of this manuscript. VFL, MK, SF, TZ, CS, EF, ABK, and TK were responsible for critical feedback and final revisions of the manuscript.

**Conflicts of Interest**

CB, VFL, SF, FW, TZ, CS, EF, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zürich, and the Institute of Technology Management at the University of St Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also the cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies were involved in any way in the design, interpretation, analysis, or writing. All other authors declare no other conflicts of interest.

**Multimedia Appendix 1**

Original (German) and translated version of the conversation flow of the hypoglycemia voice warning in study 1 and study 2. [PDF File (Adobe PDF File), 554 KB - humanfactors_v11i1e42823_app1.pdf]

**Multimedia Appendix 2**

Questions used in the semistructured interview about participants’ experience with the warning conducted in study 1 and study 2. [PDF File (Adobe PDF File), 209 KB - humanfactors_v11i1e42823_app2.pdf]

**Multimedia Appendix 3**

Illustration of the procedure across studies. [PDF File (Adobe PDF File), 174 KB - humanfactors_v11i1e42823_app3.pdf]

**References**


35. Chung AE, Griffin AC, Selezevna D, Gotz D. Health and fitness apps for hands-free voice-activated assistants: content analysis. JMIR Mhealth Uhealth 2018 Sep 24;6(9):e174 [FREE full text] [doi: 10.2196/mhealth.9705] [Medline: 30249581]

37. Dey AK, Abowd GD, Salber D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Hum Comput Interact 2001;16(2-4) [FREE Full text] [doi: 10.1207/S15327051HCI16234_02]


60. The Nightscout Foundation. URL: https://www.nightscoutfoundation.org/ [accessed 2023-12-22]


©Caterina Bérubé, Vera Franziska Lehmann, Martin Maritsch, Mathias Kraus, Stefan Feuerriegel, Felix Wortmann, Thomas Züger, Christoph Stettler, Elgar Fleisch, A Baki Kocaballi, Tobias Kowatsch. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.