

---

# JMIR Human Factors

---

Impact Factor (2023): 2.6

Volume 5 (2018), Issue 3 ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

---

## Contents

### Original Papers

Physician and Pharmacist Medication Decision-Making in the Time of Electronic Health Records: Mixed-Methods Study (e24) Kathryn Mercer, Catherine Burns, Lisa Guirguis, Jessie Chin, Maman Dogba, Lisa Dolovich, Line Guénette, Laurie Jenkins, France Légaré, Annette McKinnon, Josephine McMurray, Khrystine Waked, Kelly Grindrod. . . . .	2
A Text Message–Based Intervention Targeting Alcohol Consumption Among University Students: User Satisfaction and Acceptability Study (e23) Ulrika Müssener, Kristin Thomas, Catharina Linderöth, Matti Leijon, Marcus Bendtsen. . . . .	14
The D2Refine Platform for the Standardization of Clinical Research Study Data Dictionaries: Usability Study (e10205) Deepak Sharma, Kevin Peterson, Na Hong, Guoqian Jiang. . . . .	24
Software for Administering the National Cancer Institute’s Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events: Usability Study (e10070) Martin Schoen, Ethan Basch, Lori Hudson, Arlene Chung, Tito Mendoza, Sandra Mitchell, Diane St. Germain, Paul Baumgartner, Laura Sit, Lauren Rogak, Marwan Shouery, Eve Shalley, Bryce Reeve, Maria Fawzy, Nrupen Bhavsar, Charles Cleeland, Deborah Schrag, Amylou Dueck, Amy Abernethy. . . . .	39
Designing Emails Aimed at Increasing Family Physicians’ Use of a Web-Based Audit and Feedback Tool to Improve Cancer Screening Rates: Cocreation Process (e25) Caroline Bravo, Diego Llovet, Holly Witteman, Laura Desveaux, Justin Presseau, Marianne Saragosa, Gratianne Vaisson, Shama Umar, Jill Timmouth, Noah Ivers. . . . .	50
A Patient-Facing Diabetes Dashboard Embedded in a Patient Web Portal: Design Sprint and Usability Testing (e26) William Martinez, Anthony Threatt, S Rosenbloom, Kenneth Wallston, Gerald Hickson, Tom Elasy. . . . .	60

Original Paper

# Physician and Pharmacist Medication Decision-Making in the Time of Electronic Health Records: Mixed-Methods Study

Kathryn Mercer<sup>1</sup>, BA (Hons), MI; Catherine Burns<sup>2,3</sup>, PhD; Lisa Guirguis<sup>4</sup>, BSc Pharm, MSc, PhD; Jessie Chin<sup>5</sup>, BSc, MSc, PhD; Maman Joyce Dogba<sup>6</sup>, MD, PhD; Lisa Dolovich<sup>7</sup>, BSc Pharm, PharmD, MSc; Line Guénette<sup>8,9</sup>, BSc Pharm, MSc, PhD; Laurie Jenkins<sup>10</sup>, RN, BSc, MBA; France Légaré<sup>6</sup>, BSc, MD, MSc, PhD; Annette McKinnon<sup>11</sup>; Josephine McMurray<sup>12</sup>, MBA, PhD; Khrystine Waked<sup>1</sup>, BSc; Kelly A Grindrod<sup>1</sup>, BSc Pharm, PharmD, MSc

<sup>1</sup>School of Pharmacy, University of Waterloo, Waterloo, ON, Canada

<sup>2</sup>Centre for Bioengineering and Biotechnology, University of Waterloo, Waterloo, ON, Canada

<sup>3</sup>Systems Design Engineering, Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada

<sup>4</sup>Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada

<sup>5</sup>Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL, United States

<sup>6</sup>Faculty of Medicine, Université Laval, Ville de Québec, QC, Canada

<sup>7</sup>Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada

<sup>8</sup>Population Health and Optimal Health Practices, CHU de Québec Research Centre, Université Laval, Ville de Québec, QC, Canada

<sup>9</sup>Faculté de pharmacie, Université Laval, Ville de Québec, QC, Canada

<sup>10</sup>Healthcare Navigators, Toronto, ON, Canada

<sup>11</sup>Patient Partner, Toronto, ON, Canada

<sup>12</sup>Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, ON, Canada

**Corresponding Author:**

Kelly A Grindrod, BSc Pharm, PharmD, MSc

School of Pharmacy

University of Waterloo

200 University Avenue West

Waterloo, ON, N2L 3G1

Canada

Phone: 1 888 4567 ext 21358

Email: [kgrindrod@uwaterloo.ca](mailto:kgrindrod@uwaterloo.ca)

## Abstract

**Background:** Primary care needs to be patient-centered, integrated, and interprofessional to help patients with complex needs manage the burden of medication-related problems. Considering the growing problem of polypharmacy, increasing attention has been paid to how and when medication-related decisions should be coordinated across multidisciplinary care teams. Improved knowledge on how integrated electronic health records (EHRs) can support interprofessional shared decision-making for medication therapy management is necessary to continue improving patient care.

**Objective:** The objective of our study was to examine how physicians and pharmacists understand and communicate patient-focused medication information with each other and how this knowledge can influence the design of EHRs.

**Methods:** This study is part of a broader cross-Canada study between patients and health care providers around how medication-related decisions are made and communicated. We visited community pharmacies, team-based primary care clinics, and independent-practice family physician clinics throughout Ontario, Nova Scotia, Alberta, and Quebec. Research assistants conducted semistructured interviews with physicians and pharmacists. A modified version of the Multidisciplinary Framework Method was used to analyze the data.

**Results:** We collected data from 19 pharmacies and 9 medical clinics and identified 6 main themes from 34 health care professionals. First, Interprofessional Shared Decision-Making was not occurring and clinicians made decisions based on their understanding of the patient. Physicians and pharmacists reported indirect Communication, incomplete Information specifically missing insight into indication and adherence, and misaligned Processes of Care that were further compounded by EHRs that are not designed to facilitate collaboration. Scope of Practice examined professional and workplace boundaries for pharmacists and

physicians that were internally and externally imposed. Physicians decided on the degree of the Physician-Pharmacist Relationship, often predicated by colocation.

**Conclusions:** We observed limited communication and collaboration between primary care providers and pharmacists when managing medications. Pharmacists were missing key information around reason for use, and physicians required accurate information around adherence. EHRs are a potential tool to help clinicians communicate information to resolve this issue. EHRs need to be designed to facilitate interprofessional medication management so that pharmacists and physicians can move beyond task-based work toward a collaborative approach.

(*JMIR Hum Factors* 2018;5(3):e24) doi:[10.2196/humanfactors.9891](https://doi.org/10.2196/humanfactors.9891)

## KEYWORDS

shared decision-making; electronic health records; collaboration; interprofessional collaboration; medication management

## Introduction

In clinical settings, medication-related decisions are often passed verbally among patients, doctors, nurses, and pharmacists, and the message can become distorted. Too often, however, critical information is not shared, even when an electronic health record (EHR) is used, and the decision to prescribe or not prescribe, to take or not take a medication, is made with missing or distorted information [1-4]. Health systems now promote an ethos of partnership where providers and patients navigate complex relationships and interactions. The shift from a patient-physician decision-making dyad to a network of providers introduces more complexity into what are often byzantine processes that precede health decisions. Nevertheless, patients often rely on a trusted health care professional's (HCP's) expertise to make important decisions where the situation is emergent or ambiguous (eg, having a surgery or starting a new medication) [5,6]. Research has not yet empirically characterized how current communication between health care practitioners affects care, and specifically how EHRs can strengthen communication by making information easier to access [7].

A medication-related decision involves, at minimum, a patient, a prescriber, and a pharmacist, and all parties are engaged in a process of shared decision-making (SDM) [8,9]. SDM is based on a model of communication where HCPs and patients both contribute to clinical decisions in unique ways [10,11]. The HCPs share information about the benefits and risks of different treatment options; the patients describe their preferences and values as they relate to their treatment options. Interprofessional shared decision-making (IP-SDM) involves multiple HCPs and is emerging as a response to care increasingly being delivered by interprofessional teams to collaboratively work with a patient to decide on the best course of action [12]. A systematic review of the adoption of SDM by HCPs concluded that while it is unclear whether interventions that promote the adoption of SDM are effective, interventions that target patients and HCPs simultaneously are more effective than ones that only target one group [13]. The evolution of IP-SDM is challenging our beliefs about how and when HCPs actively communicate with each other and with patients as well as about the role EHRs may play in decision-making.

Adverse drug events (ADEs) are one of the outcomes of miscommunication in the medication management process. The costs of ADEs to the health care system are staggering, yet in

one US study, physician reviewers determined that of the 30% inpatients who experienced ADEs, the events were preventable in 44% cases [14-16]. While these medication-related problems are the symptom of a complex and disconnect health care system, the inclusion of pharmacists in medication management has reduced the rates of ADEs as well as health care costs [17]. ADEs account for somewhere between 1.4% and 15.4% of hospital admissions in the United States and Canada, accounting for an estimated 177,504 emergency department visits by US patients aged  $\geq 65$  years and increasing the mean length of hospital stay from 8 to 20 days [18-20]. SDM is known to improve communication, lessen ADEs, and, overall, lower health care costs [21,22]. Through greater communication and collaboration between HCPs and patients, IP-SDM provides a platform that has significant potential to further lessen ADEs and to continue lower health care costs [23].

In most health care settings, pharmacists and physicians often do not communicate well because they largely work *independently and in parallel* with each other, rather than collaboratively [24]. Furthermore, there can be challenges in communication due to differences of opinion on role, reluctance to challenge, different work schedules, and different information priorities [25-27]. For example, how physicians and pharmacists communicate and make decisions with each other is based on perceptions about the role each one plays in a person's care and is tightly tied to ideas about pharmacists' scope of practice. According to Nugus et al, there is a clear acknowledgment in health care that physicians are the ones with "formal responsibility of patient care" and that they are omnipresent in care [28]. As a result, EHRs may reflect the physician's information or decision-making needs more than those of the pharmacist or the patient. The challenge in designing interdisciplinary EHRs is that they need to account for the workflow and communication models of different professions. It is important that physicians and pharmacists have a good communication because it is essential to go beyond transactional interactions to ensure optimal therapeutic outcomes for patients [29]. This research has been conducted to better foundationally understand how pharmacists and physicians communicate, which can be used to lessen medication-related errors, lower health care costs, and design and improve EHRs that facilitate collaboration.

The objective of this exploratory study is to examine how physicians and pharmacists understand and communicate patient-focused medication information with each other and to

identify barriers to IP-SDM for medication management that should inform designing EHRs that support IP-SDM. This research will allow for the design and refinement of EHRs that can be designed to facilitate better communication, improve medication management, and ultimately contribute to improved care.

## Methods

### Research Design

This research was part of a larger mixed methods study on SDM in the context of EHRs that included observations; interviews; and think-aloud discussions with patients, primary care physicians, and pharmacists. This paper focuses on qualitative, semistructured interviews with physicians and pharmacists. We have taken a pragmatic stance, recognizing that a constructivist view of the truth can be tempered with the need to conduct research that informs health care decision-making [30]. Our analysis was guided by a framework analysis method that provides both a systemic and flexible approach to multidisciplinary data analysis [31].

We conducted interviews in community pharmacies and primary care clinics across Canada using provinces to represent different levels of primary care integration and adoption of EHRs (Table 1). This research received ethics approvals from the University of Waterloo, the University of Alberta, Wilfrid Laurier

University, Université Laval, the University of Toronto, and Dalhousie University.

### Recruitment and Participants

The research team used a purposive sampling approach to identify a broad spectrum of practice sites. Recruitment was conducted through several venues including posters, social media, and snowball sampling from previous and existing contacts of the research team. We included pharmacists and family physicians practicing in Ontario, Alberta, Quebec, and Nova Scotia.

### Data Collection

Three research assistants conducted and audiorecorded the interviews. One of the research assistants was a PhD candidate and experienced qualitative researcher (KM) and the other two were PharmD students (KW, JB). The three interviewers jointly conducted 3 interviews to train the student research assistants in semistructured interview techniques, and they regularly met throughout the data collection period to compare interview notes and transcripts. All three research assistants interviewed participants in Ontario, with KW completing all of the interviews in Quebec and Alberta and JB completing all of the interviews in Nova Scotia. Field notes recorded during and after the interviews documented the environment, external influencers or distractions, and participants; specific questions were added to better understand the decision-making approach.

**Table 1.** Description of in-place electronic health records (EHRs) and primary care models in Alberta, Ontario, Quebec, and Nova Scotia between December 2015 and October 2016.

Health record	Province			
	Alberta	Nova Scotia	Ontario	Quebec
<b>EHR<sup>a</sup></b>	Netcare	SHARE <sup>b</sup>	ClinicalConnect <sup>c</sup>	Dossier Santé Québec
Medication profile	Yes	Drug Information System	Only hospital medications	Yes
Laboratory values	Yes	Yes	Yes	Yes
Medical imaging	Yes	Yes	Yes	Yes
Integrated systems	Pharmaceutical Information Network	Drug Information System	Ontario Laboratories Information System	N/A <sup>d</sup>
Other information	Hospital visits, surgeries, drug alerts, allergies or intolerances, immunizations	Hospital admissions or discharge information, history and consulting notes	Allergies, medical reports, pathology and microbiology results	Electronic prescriptions
Physician access to EHR	Yes	Yes	Yes	Yes
Pharmacist access to EHR	Yes	Drug Information System <sup>e</sup>	No	Yes
Team-based health care	Primary Care Network health care teams	Collaborative care teams	Family health teams	Family medicine groups
Pharmacist integration in team-based health care	Yes	Yes	Yes	Government promotes close ties between community pharmacies and family medicine groups

<sup>a</sup>Information collected in this table reflects health care at the time of the interviews and may have changed since.

<sup>b</sup>SHARE: Secure Health Access Record.

<sup>c</sup>South Western ON. EHRs are region specific in Ontario; separated into 3 regions.

<sup>d</sup>N/A: not applicable.

<sup>e</sup>Access to laboratory values in near future.

Interviews with HCPs consisted of two parts: (1) medication-focused decision-making and (2) interviewee's opinion of EHRs. HCPs were interviewed where they practiced, either in the pharmacy or the physician's office. Interviews focused on how the pharmacist or physician presented information to patients; how collaboration was approached during care, specifically in relation to medication prescribing or problem solving; how they interacted with EHRs or electronic medical records (EMRs) used in their practice; and finally potential areas for developing new EHRs. The interview guide is available in [Multimedia Appendix 1](#).

### Data Analysis

We employed a modified version of the *Multidisciplinary Framework Method* to analyze the data [24]. A multidisciplinary team, including engineers, clinicians, health researchers, business and communication researchers, patients, and a patient navigator, was involved in data analysis. The steps followed were as follows: (1) interviews were transcribed verbatim; (2) core research team members read the transcripts and listened to the audiorecordings to familiarize themselves with the interviews; (3) core team members thematically coded the data; (4) the entire team thematically coded a subset of 5 interviews; (5) the team codes were used to develop a working analytic framework; (6) 2 team members (KM, KW) recoded the data; and (7) finally, the data were presented to the entire team for discussion and refinement. Data were stored, organized, and

reported using QSR NVIVO 11 Software (QSR International Pty Ltd. Version 11, 2017). Any names and identifiers were made anonymous in the transcription process. Multiple triangulation of the data was achieved using a variety of geographic sources, multiple coders, and a multidisciplinary team of researchers interpreting the results [32].

## Results

### Study Population

In total, we interviewed 25 pharmacists and 9 family physicians (Table 2). On average, the HCPs had been with their current clinic for 8 years and had been practicing for 15 years. Compared with physicians, a larger sample of pharmacists was recruited to account for variability in practice setting; the latter included pharmacists who worked in chain pharmacies (n=5), independently owned pharmacies (n=12), and team-based medical clinics (n=4).

### Thematic Analysis

Initial coding conducted by the core research team led to the identification of 46 codes, which were then developed into 5 themes describing the different elements of how pharmacists and physicians make medication-related decisions with patients: *workflow*, *communication*, *accuracy*, *decision-making*, and *computer systems*.

**Table 2.** Participant demographics collected at the time of interview (N=34).

Characteristics	Family physicians	Pharmacists
<b>Province, n</b>		
Nova Scotia	0	4
Quebec	2	2
Ontario	6	15
Alberta	1	4
Total participants, n	9	25
Team environment, n	9	4
Independent practice, n	0	21
Years in practice <sup>a</sup>	12.6	16.2
Average duration of current practice (years)	9.9	7.1
Average age of participants (years) <sup>a</sup>	43.4	39.8
<b>Participant age group (years), n<sup>a</sup></b>		
25-35	2	7
36-45	4	12
46-55	2	4
55+	1	2
<b>Gender</b>		
Male	4	11
Female	7	14

<sup>a</sup>Information regarding age and years in practice was not collected from 1 family physician participant.

As part of the multidisciplinary framework, we held a 2-day research meeting where the entire multidisciplinary team participated in the analysis. Research group members came to the meeting having individually coded the same 5 interviews. Through a process of negotiation, individual codes were rearranged into 81 subthemes and 6 major themes as outlined below (Table 3). KM and KW recoded the remaining interviews using the new framework, with no additional themes arising.

The new coding framework placed a more significant focus on how pharmacist-physician relationships and scopes of practice affect medication-related decisions (Table 3). We found that decision-making was influenced by the information, processes, and communication factors related to EHRs, which, in turn, were influenced by the physician-pharmacist relationships and scopes of practice.

**Table 3.** Themes related to interprofessional medication-related decision-making between physicians and pharmacists.

Theme	Subthemes	Description
Interprofessional shared decision-making (IP-SDM)	<ul style="list-style-type: none"> <li>IP-SDM intentions</li> <li>Decision point</li> <li>Making the decision</li> <li>Assumptions about patients</li> <li>Patient communication</li> <li>IP-SDM</li> </ul>	Pharmacists and physicians did not describe IP-SDM in their practices and acted as unintentional gatekeepers to medication information. Professionals make decisions based on their individual understanding of the patient's situation and educate the patient based on that decision.
Communication between physicians and pharmacists	<ul style="list-style-type: none"> <li>Reasons for health care professionals (HCPs) to communicate with patients</li> <li>Reasons for HCPs to communicate with each other</li> <li>Flow of information</li> <li>Communication workarounds</li> <li>Method of communication</li> <li>Availability</li> <li>How to document in the medical or pharmacy chart</li> <li>Risk communication</li> <li>Patient as messenger</li> </ul>	Pharmacists and physicians often communicate with each other indirectly through patients, faxes, or receptionists. Yet, both groups are cautious about the expansion of electronic health records (EHRs) and how EHRs influence their ability to work.
Information exchange between physicians and pharmacists	<ul style="list-style-type: none"> <li>Important information for patient care</li> <li>Information detectives</li> <li>Data collection and entry</li> <li>Multiple users</li> <li>Place of access</li> <li>Context of data entry</li> <li>Adherence</li> <li>Information scarcity limits roles</li> <li>Design features</li> <li>Timeliness</li> </ul>	Pharmacists and physicians require information not accessible through current Web-based health platforms to provide patient care. Even in situations where the information was available, it was clear that relationships drove information sharing. Most critically, physicians required access to information about medication adherence, while pharmacists required clear access to medication indications.
Process of care	<ul style="list-style-type: none"> <li>System design (fill and bill)</li> <li>Identifying patients in need of care</li> <li>Stages of care</li> <li>Technology limits practice</li> <li>Decision-making</li> <li>Workarounds</li> <li>Documentation of process</li> <li>Workflow bottlenecks</li> <li>Prioritization</li> </ul>	Pharmacists and physicians find that current systems do not typically align with their decision-making processes and do not support collaboration in daily workflow.
Scope of practice	<ul style="list-style-type: none"> <li>Responsibility to diagnose</li> <li>Negotiating role boundaries</li> <li>Accountability</li> <li>Medication management</li> <li>Mentorship and role modeling</li> <li>Monitoring</li> </ul>	The workplace and professional boundaries for pharmacists and physicians are both internally and externally imposed. This includes how each group negotiates the boundaries of its job, how each group negotiates its interactions with each other and with patients, and how relationships, or lack thereof, impact the ability of each group to carry out its roles and responsibilities.
Physician-pharmacist relationship	<ul style="list-style-type: none"> <li>Physical distance</li> <li>Community versus primary care pharmacist</li> <li>5 Ws of shared understanding</li> <li>Filling the gap or tailoring</li> <li>Building collaborative work environments</li> <li>Transactional communication</li> </ul>	Relationships were strongly influenced by physician location, nature of the task, and a power imbalance.

## Interprofessional Decision-Making

In the interviews, we asked about how different treatment options were presented, how patients' values were taken into account, and whether the participant knew about IP-SDM. We observed that IP-SDM was not an active part of the typical decision-making process. Rather, we identified a spectrum of decision-making, where the most common approaches to decision-making included paternalism and informed decision-making, as outlined below, rather than IP-SDM.

In the paternalistic decisions that were both described and witnessed, the physician or pharmacist made a decision because they "assumed," "understood," or "knew" it was the "best", and then, they "informed" the patient regarding what the patient should do. In other words, the physicians or pharmacists "shared" *their* final decision rather than sharing the decision-making process:

*I really do consider also the patient's preference or pre-knowledge or understanding. Have I considered all the factors; the patient factors, cost factors? That kind of thing I try to make it so it's sort of like rational prescribing, thinking is there a reason to give it to them?* [Physician 1205, Family Health Team, Ontario]

During informed decision-making, pharmacists and physicians focus on educating patients well enough to allow them to make a decision. The goal is to offer recommendations, help the patient understand why the HCP offered the recommendation, and allow the patient to choose whether he or she wants to pursue the recommended course of action:

*I want them to make an informed decision. I want them to understand what's going on with their health. I want them to understand what the options are and why we're pursuing those options. I want them to make an informed decision about whether they want to move forward with a particular treatment course or not and understand the rationale for that.* [Physician 1202, Family Health Team, Ontario]

One of the challenges of informed decision-making is that the information could "scare" the patient. It is unrealistic for all patients to become as well educated as an HCP about a medical decision:

*I don't want to give more information than necessary, especially if I see that a patient is more anxious during the beginning of the counselling, and even more so if the patient doesn't want to take the medication or is scared to take the medication.* [Pharmacist 1121, Quebec, Independent Pharmacy]

Pharmacists who worked in teams talked of making decisions *with* physicians rather than patients:

*It was last Wednesday, was the last day that I worked there, and it was more I help the physician choose the medication. Not so much the patient themselves. It was a very complex case and the physician had asked me to meet with the patient first to do a medication review appointment.* [Pharmacist 1124, Family Health Team Ontario]

## Communication Between Pharmacists and Physicians

Communication between pharmacists and physicians is heavily dependent on the fax machine. Unlike a phone, faxed documents provide a written record of an encounter. However, fax machines are not connected with pharmacist and physician information systems, reducing the efficiency of their use.

*We almost prefer a fax than phone a physician. We phone if it's an immediate thing, but faxing gives us, again, the detailed paper, dated and detailed work that we can keep track of. That's what we try to do* [Pharmacist 1109, Independent Clinic, Nova Scotia]

A common complaint among participants was that the standard processes to request information from another HCP are flawed. Pharmacists felt that they were limited by having to wait for a reply to a fax, and physicians often mentioned waiting until they had time to track down a pharmacist they trusted. The notion of a centralized way to communicate information was met with positive reactions. Being able to access key information without actively and asynchronously communicating with another HCP was identified as a way to streamline the sharing of basic medical information (eg, diagnosis, prescriptions, and lab results). Communication might then be focused around sharing meaningful information, such as patient histories or complex care regimens. Participants were concerned that information is not properly being communicated and may be missing or incorrectly documented. Pharmacists reported rarely being able to get past gatekeepers, such as office staff.

*There's the ward clerk who won't let you through to the doctor. It's really difficult to get a doctor on the phone unless if they're calling you* [Pharmacist 1102, Independent Pharmacy, Ontario]

At instances where pharmacists are colocated with physicians, face-to-face interactions have the potential to foster the development of a trusting relationship. However, even when pharmacists and physicians are colocated, pharmacists still negotiate the power differential by modulating communication.

*I don't go to a physician and say, "You must do this." I say, "This is the problem that this patient is having on these medications. Here are our options. The options are A, B, and C. I like A because this, this, and this. I like B because of this, this, and this. What do you think we should do?" I never try and tell them what to do.* [Pharmacist 1125, Family Health Team, Ontario]

## Information Exchange Between Pharmacists and Physicians

Pharmacists and physicians use different pieces of information to provide patient care. Physicians record diagnostic information, including physical evaluations and tests, while pharmacists keep detailed records of medications provided. Most of the interviewed community pharmacists did not have access to the reason a medication was prescribed, the diagnostic test, or laboratory results. They assessed appropriateness and dispensed medications using the limited information contained on a prescription or patient recall. Additional or clarifying information needed to be requested from the physician. Even

in situations where pharmacists had access to information through an EHR, issues related to missing information, and the consequent need to contact a physician to gain access to it, were mentioned.

*Maybe there's some piece of information that we're missing and that's where you ask questions. If they're asking for refills too soon then it may be, "Why are you needing this more than what has been prescribed? Are you taking more than what was on the instructions that we have? Has someone told you to take more?"*  
[Pharmacist 1124, Family Health Team, Ontario]

Pharmacists often were missing information on the reason medications were prescribed. Not knowing why or how a prescriber decided on a medication not only limited pharmacists' ability to properly educate patients about their medications but also limited their ability to participate in decisions to start, change, or stop a medication.

*I would just say that getting information on the indication would be one. Trying to find out what they're taking the medication for and what they're hoping it's going to do for them would be two of the first questions.* [Pharmacist 1107, Nova Scotia, Independent Practice]

Physicians were also concerned that pharmacists do not have sufficient patient information to effectively prescribe, deprescribe, or assess a patient's medications. Physicians were missing information about how a medication is taken after it was prescribed. Occasionally, there were plans in place to confirm prescription pick up and adherence, but the absence of adherence data was a clear gap in information. Both groups cited that the benefit of an EHR was enabling improved communication and overall patient care.

*We've got a system [to help us keep track of] adherence. It's a really difficult point, and it's a really important point that I think we need to look because it's not good right now.* [Physician 1201, Ontario, Family Health Team]

### Process of Care

Pharmacists and physicians follow different processes for providing care, which are reflected in different information systems used in their daily workflow. Physicians use clinical data from physical assessment, lab values, and diagnostic imaging to make treatment decisions. Their office-based EMRs support documentation of their patient encounters, assessments of the information, and prescribing history. In community pharmacies, pharmacy practice management systems support dispensing and provide patient information sheets, auxiliary warning labels, warnings about adherence, and drug interaction alerts. Even in Alberta and Quebec, where there are province-wide EHRs that include lab values and dispensing information, the dispensing information is not integrated into the physician's EMR and the clinical information is not incorporated into the pharmacist's pharmacy practice management systems.

*My goal is to get my EMR and the pharmacist's EMR exactly the same and up to date.* [Physician 1206, Family Health Team, Alberta]

The lack of overlap between physician and pharmacist information systems reinforces the siloed workflows of the two professions and lack of interoperability between privately-owned EMRs. However, even when pharmacists and physicians work on the same system, it can be difficult to mesh the two decision-making processes. The resulting hybrid can be inefficient, requiring back-and-forth between the patient and different HCPs.

*I made some recommendations to the physician and patient, which then the physician discussed with the patient in her appointment with the patient. We also discussed, the doctor and I, after, to confirm, yes, this is what we did, and just to follow-up on the whole discussion.* [Pharmacist 1124, Family Health Team, Ontario]

Many participants lacked awareness of the decision-making processes of other HCPs, which left them guessing why certain decisions were made. Guesswork, thus, becomes the *de facto* process, rather than an open and collaborative process. Finally, even though Alberta pharmacists are able to prescribe and use a provincial EHR used by physicians, their experiences have been ultimately similar to the pharmacists in other provinces who did not have access to an EHR.

### Scope of Practice

Scope of practice refers to the internal and external boundaries placed on pharmacists and physicians. In many provinces, the scope of pharmacist practice has expanded to include prescribing, which has traditionally been the physician's role. This can result in role friction.

*It's been good, all the changes, for sure. [...] You just learn [which physicians] who you can do it with and who you can't, and then you go with that.* [Pharmacist 1114, Independent Pharmacy, Alberta]

In rural areas, pharmacists have more latitude toward full scope of practice as fewer options for care are available, and they are more likely to know other local HCPs. Scarcity of services provides situations that encourage greater collaboration and partnerships due to availability as well as familiarity with colleagues.

*There's no full time physician in town...A lot of the local doctors are very open to our input and actually will seek it. Nearby doctors are a group who will cover for each other, and we know them.* [Pharmacist 1110, Independent Pharmacy, Nova Scotia]

Ideally, a team-based practice means that the different professions are more easily able to understand each other's roles, including how one profession's skills can complement another's. Physicians generally did not consider pharmacists as partners in care and rarely mentioned active collaboration.

*Yeah. Things are good with my pharmacist and I. We're still trying to work on enhancing our relationship but definitely the trust exists there and*



*then now it's just kind of more a matter of allowing some pharmacists to feel like they can do more.*  
[Physician 1205, Family Health Team, Ontario]

Even in cases where active collaboration was spoken of in a meaningful and positive way, it was still clear that there were underlying restrictions; for example, in the above quote, while the physician spoke about collaboration, the comments qualified that only *some* pharmacists should be allowed to *feel* like they could do more. Similarly, the physician referred to the pharmacist team member as “my pharmacist,” creating in and out groups of pharmacists and reinforcing traditional power archetypes.

### Relationships Between Pharmacists and Physicians

Physician-pharmacist relationships were often influenced by physical location and institutional context. When pharmacists and physicians were colocated, particularly when a common institutional governance was present, such as a family health team in Ontario, they were able to share a common system of health records. The face-to-face interactions also allowed the pharmacists and physicians to establish personal relationships with each other. Building trusting relationships allowed for informal collaboration about patient care. Pharmacists often spoke of feeling like an outsider to care or that they were “... not wanting to bother” the physicians [Pharmacist 1107, 1108, 1109, 1121]. The limited opportunity for face-to-face collaboration artificially restricted the pharmacist’s ability to support the patient.

Pharmacists also often felt that they had to navigate the authority of physicians when assessing medication, and that, due to their perceived role in health, they were not able to influence care to the best of their abilities.

*I notified a patient’s physician to a contraindicated drug given by a patient’s psychiatrist. The physician didn’t feel comfortable changing the drug, and the psychiatrist said, well, I’m not changing mine, I have him on what I want him to be on. The neurologist, I couldn’t get in touch with him, and then the group home, they were almost a little bit, “we wish you hadn’t put your hands in the pot, there’s too many people trying to mess things up.” It was really frustrating because there’s this clear thing that could cause harm to the patient, and you almost felt like you were doing more harm than good by alerting everyone to it.* [Pharmacist 1102, Independent Pharmacy, Ontario]

Finally, the interviews made it clear that pharmacists’ processes for working with physicians are not designed to facilitate collaboration. Rather, they may have evolved as workarounds that compensate for the strained relationship with the physician.

*Most physicians do like subtle language of requesting as to, “Can you give me the thought behind prescribing this because we’re just not sure, we want to make sure the patient understands it well or providing recommendations.”* [Pharmacist 1116, Alberta, Chain Pharmacy]

## Discussion

### Principal Findings

This project examines how physicians and pharmacists communicate patient-focused medication information with each other to inform the design of EHRs for IP-SDM. There is limited research on how EHRs currently impact IP-SDM and the potential they have for improving collaboration. We can see that the limited communication between physicians and pharmacists is strongly dependent on the relationship between them. The suboptimal management and use of medication have already been well documented, suggesting that we may not be optimally positioned to provide accessible, effective, and affordable medication management as patient need rises over the coming decade [33]. Before pharmacists and physicians can share medication-related decisions with patients, they themselves need access to comprehensive information. Furthermore, they must be prepared to share information about decision-making and to develop strategies for interprofessional collaboration that do not rely on colocation or a common institutional EMR or EHR. The findings of this study point to a status quo where integrated provider medication management and IP-SDM are an exception rather than the rule in community settings.

Workable solutions to how information is shared are both social and technical. Most electronic health information systems are capable of semantic interoperability, where a receiving information system is able to clearly interpret information in exactly the same way as the sending information system. Use of vocabularies, including RxNorm, and structured documents, such as the Clinical Document Architecture and Fast Healthcare Interoperability Resources, supports interoperability [34]. As beneficial as these may be, the competitive market forces the costs that rarely support this option, despite its popularity among providers. Despite pharmacists having played an integral role in delivering high-quality clinical care in hospitals for decades, this study highlights the slow progress toward integration and IP-SDM acceptance in the community. Our research supports the idea that social factors such as professional acceptance, institutional structures, and trusting versus nontrusting relationships are more significant barriers to the adoption of EHRs into patient care compared with technical challenges.

Kannampallil et al [35] have noted that “complex systems can appear very different, depending on the aspects, granularity, and circumstances that the researcher chooses to focus on.” By focusing on the relationship between physicians and pharmacists in this study, we saw that each health care profession has access to critical information that the other profession does not (eg, pharmacists do not have access to information about a medication’s reason for use and physicians do not have access to adherence information). These reasons are related to inadequate systems for health information exchange as well as missing professional standards that encourage comprehensive medication information exchange.

Our findings on communication, information, and process mirrored Bardet et al’s meta-model on physician and community pharmacist collaboration [36]; they identified that early on in a collaboration, key elements include trustworthiness and clarity

around roles. Physicians and pharmacists also need to develop an interdependence; establish interest, skills, and positive perceptions; have clear expectations; and build a relationship that is grounded in trust [37,38]. Open and bidirectional communication is also important [36]. Our findings add to the work of Bardet et al by highlighting how the disconnected computer systems and decision processes limit collaboration between pharmacists and physicians. All participants were enthusiastic about the potential for provincial EHRs to improve information sharing and communication [39]. A well-designed EHR could also facilitate many components of a successful collaboration. Specifically, it has the potential to foster IP-SDM and level the playing field for understanding around information, process, and communication.

According to a review of IP-SDM by Dogba et al, safe and high-quality health care depends on increased levels of collaboration among HCPs and better engagement with patients [40]. In our study, all participants voiced their support for IP-SDM in general. However, when it came to giving examples, only one physician was able to describe an instance of IP-SDM in practice, and no pharmacists or physicians were able to clearly articulate a shared vision for IP-SDM. Moreover, participants had reservations about their patients' ability to make decisions. They referenced the notion that HCP training and experience enable them to know what is "best for the patients." Patel et al [41] have referred to this as a "cautious willingness" to participate in IP-SDM due to fears over patient competence, motivations, and dishonesty about adherence.

The notion of "cautious willingness" also applies to HCP collaboration [42]. Physicians are cautious about giving up a perceived ownership of a patient's care, and pharmacists are equally cautious about making physicians feel like they are trying to take over the care. The reluctance of pharmacists to embrace a full scope of practice also reflects serious concerns about missing information. In the interviews, it was clear that pharmacists perceive themselves as the last gatekeeper of a patient's well-being, yet they are unable to perform that function.

Elwyn et al [43] noted that HCPs often miss the second half of a consultation, where IP-SDM occurs. We would argue that the second half of the medication-related consultation is where IP-SDM and the pharmacist belong. Physicians have the unique expertise to focus on the diagnoses in the first half of the consultation. Pharmacists, however, have the expertise required to help the patients understand and choose a treatment option that is consistent with their needs and preferences. However, pharmacists cannot act until they have access to the right information at the right time and have a bidirectional

communication with the physician. Ultimately, research should evaluate the link between all interactions in the health care process that impact patient and clinician decision-making.

### Strengths and Limitations

As part of a larger mixed methods study, the insights presented here are derived solely from the interviews of pharmacists and physicians. Although these analyses reveal perceptions about and barriers to IP-SDM and collaboration, they do not reflect a complete analysis of all data collected, specifically the data collected from patients. However, in the context of gaining a deep understanding of physician-pharmacist communications and relationships, this analysis is a critical step in building a holistic model of IP-SDM related to medication management. In addition, while the sample includes pharmacists across all 4 provinces, recruitment challenges limited the participation of physicians in each of the 4 provinces, especially in Nova Scotia. Given the similarities in policies and practice across Canadian provinces and the inclusion of a variety of physician perspectives, we believe this had little to no impact on our results. Finally, differences in interviewers' approaches to semistructured interviews may have led to differing emphasis on IP-SDM and collaboration. While the benefit of a multidisciplinary research team is stronger objectivity stemming from a variety of research, professional, and patient backgrounds, this study might have been strengthened even more if the research team had employed prolonged engagement. Although important, due to interview time constraints, we could not explore physicians' perceptions of pharmacists prescribing, adapting, or cancelling medications; the influence of these perceptions is suggested to be explored in future research.

### Conclusion

Our study shows that until pharmacists can see the reason for which a medication is prescribed and physicians gain insight into adherence, neither group will be fully able to work together to make medication-related decisions collaboratively. The major barriers to collaboration include poor communication systems with minimal interinstitutional information exchange, and even when an EHR exists, competing decision-making processes are most often present. We identified the potential to build EHRs that not only better facilitate access to information but also allow for processes that better accommodate collaborative care and enable better understanding of the pharmacist's scope of practice. Future research should focus on the alignment of EHRs with interprofessional decision-making process, which can foster both intra- and interinstitutional collaboration and information sharing to best support IP-SDM.

---

### Acknowledgments

We would like to thank Christian Chabot for his support and input throughout the project and Jonathan Boersema for assisting with data collection.

---

### Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Interview Guide.

[[PDF File \(Adobe PDF File\), 140KB - humanfactors\\_v5i3e24\\_app1.pdf](#)]

### References

1. Hall AE, Paul C, Bryant J, Lynagh MC, Rowlings P, Enjeti A, et al. To adhere or not to adhere: Rates and reasons of medication adherence in hematological cancer patients. *Crit Rev Oncol Hematol* 2016 Jan;97:247-262. [doi: [10.1016/j.critrevonc.2015.08.025](#)] [Medline: [26412718](#)]
2. Polinski JM, Kesselheim AS, Frolkis JP, Wescott P, Allen-Coleman C, Fischer MA. A matter of trust: patient barriers to primary medication adherence. *Health Education Research* 2014 May 16;29(5):755-763. [doi: [10.1093/her/cyu023](#)] [Medline: [24838119](#)]
3. Greenhill N, Anderson C, Avery A, Pilnick A. Analysis of pharmacist-patient communication using the Calgary-Cambridge guide. *Patient Educ Couns* 2011 Jun;83(3):423-431. [doi: [10.1016/j.pec.2011.04.036](#)] [Medline: [21640542](#)]
4. van Dijk M, Blom L, Koopman L, Philbert D, Koster E, Bouvy M, et al. Patient-provider communication about medication use at the community pharmacy counter. *Int J Pharm Pract* 2016 Feb;24(1):13-21. [doi: [10.1111/ijpp.12198](#)] [Medline: [25988371](#)]
5. Becker ER, Roblin DW. Translating primary care practice climate into patient activation: the role of patient trust in physician. *Med Care* 2008 Aug;46(8):795-805. [doi: [10.1097/MLR.0b013e31817919c0](#)] [Medline: [18665059](#)]
6. Müller E, Zill JM, Dirmaier J, Härter M, Scholl I. Assessment of trust in physician: a systematic review of measures. *PLoS One* 2014 Sep;9(9):e106844 [FREE Full text] [doi: [10.1371/journal.pone.0106844](#)] [Medline: [25208074](#)]
7. Tharmalingam S, Hagens S, Zelmer J. The value of connected health information: perceptions of electronic health record users in Canada. *BMC Med Inform Decis Mak* 2016 Jul 16;16:93 [FREE Full text] [doi: [10.1186/s12911-016-0330-3](#)] [Medline: [27422571](#)]
8. Pande S, Hiller JE, Nkansah N, Bero L. The effect of pharmacist-provided non-dispensing services on patient outcomes, health service utilisation and costs in low- and middle-income countries. *Cochrane Database Syst Rev* 2013 Feb 28(2):CD010398. [doi: [10.1002/14651858.CD010398](#)] [Medline: [23450614](#)]
9. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. *Social Science & Medicine* 1999 Sep;49(5):651-661. [doi: [10.1016/S0277-9536\(99\)00145-8](#)]
10. Lin GA, Fagerlin A. Shared Decision Making: State of the Science. *Circulation: Cardiovascular Quality and Outcomes* 2014 Feb 04;7(2):328-334. [doi: [10.1161/CIRCOUTCOMES.113.000322](#)] [Medline: [24496297](#)]
11. Elwyn G, Lloyd A, May C, van der Weijden T, Stiggelbout A, Edwards A, et al. Collaborative deliberation: A model for patient care. *Patient Education and Counseling* 2014 Nov;97(2):158-164. [doi: [10.1016/j.pec.2014.07.027](#)] [Medline: [25175366](#)]
12. Légaré F, Stacey D, Gagnon S, Dunn S, Pluye P, Frosch D, et al. Validating a conceptual model for an inter-professional approach to shared decision making: a mixed methods study. *J Eval Clin Pract* 2011 Aug;17(4):554-564 [FREE Full text] [doi: [10.1111/j.1365-2753.2010.01515.x](#)] [Medline: [20695950](#)]
13. Légaré F, Stacey D, Turcotte S, Cossi M, Kryworuchko J, Graham ID, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. *Cochrane Database Syst Rev* 2014 Sep 15(9):CD006732. [doi: [10.1002/14651858.CD006732.pub3](#)] [Medline: [25222632](#)]
14. Leendertse AJ, Van Den Bemt PMLA, Poolman JB, Stoker LJ, Egberts ACG, Postma MJ. Preventable hospital admissions related to medication (HARM): cost analysis of the HARM study. *Value Health* 2011 Jan;14(1):34-40 [FREE Full text] [doi: [10.1016/j.jval.2010.10.024](#)] [Medline: [21211484](#)]
15. Hohl CM, Nosyk B, Kuramoto L, Zed PJ, Brubacher JR, Abu-Laban RB, et al. Outcomes of emergency department patients presenting with adverse drug events. *Ann Emerg Med* 2011 Sep;58(3):270-279.e4. [doi: [10.1016/j.annemergmed.2011.01.003](#)] [Medline: [21354651](#)]
16. Office of the Inspector General. Washington, DC: Department of Health and Human Services; 2010 Nov. Adverse Events in Hospital: National Incidence Among Medicare Beneficiaries URL: <https://oig.hhs.gov/oei/reports/oei-06-09-00090.pdf> [accessed 2018-08-09] [WebCite Cache ID 71XoIO5fq]
17. Butler A, Dehner M, Gates RJ, Shane P, Chu M, DeMartini L, et al. Comprehensive Medication Management programs: 2015 status in Southern California. *Res Social Adm Pharm* 2017 Jan;13(1):63-87. [doi: [10.1016/j.sapharm.2016.02.003](#)] [Medline: [27234647](#)]
18. Howard RL, Avery AJ, Slavenburg S, Royal S, Pipe G, Lucassen P, et al. Which drugs cause preventable admissions to hospital? A systematic review. *Br J Clin Pharmacol* 2007 Feb;63(2):136-147 [FREE Full text] [doi: [10.1111/j.1365-2125.2006.02698.x](#)] [Medline: [16803468](#)]
19. Budnitz DS, Shehab N, Kegler SR, Richards CL. Medication Use Leading to Emergency Department Visits for Adverse Drug Events in Older Adults. *Ann Intern Med* 2007 Dec 04;147(11):755. [doi: [10.7326/0003-4819-147-11-200712040-00006](#)]

20. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *PLoS One* 2009 Feb;4(2):e4439 [FREE Full text] [doi: [10.1371/journal.pone.0004439](https://doi.org/10.1371/journal.pone.0004439)] [Medline: [19209224](https://pubmed.ncbi.nlm.nih.gov/19209224/)]
21. Wright SM, Hedin SC, McConnell M, Burke BV, Watts SA, Leslie DM, et al. Using Shared Decision-Making to Address Possible Overtreatment in Patients at High Risk for Hypoglycemia: The Veterans Health Administration's Choosing Wisely Hypoglycemia Safety Initiative. *Clin Diabetes* 2018 Apr;36(2):120-127. [doi: [10.2337/cd17-0060](https://doi.org/10.2337/cd17-0060)] [Medline: [29686450](https://pubmed.ncbi.nlm.nih.gov/29686450/)]
22. Phansalkar S, Hoffman JM, Hurdle JF, Patel VL. Understanding pharmacist decision making for adverse drug event (ADE) detection. *J Eval Clin Pract* 2009 Apr;15(2):266-275 [FREE Full text] [doi: [10.1111/j.1365-2753.2008.00992.x](https://doi.org/10.1111/j.1365-2753.2008.00992.x)] [Medline: [19335483](https://pubmed.ncbi.nlm.nih.gov/19335483/)]
23. Yu CH, Ivers NM, Stacey D, Rezmovitz J, Telner D, Thorpe K, et al. Impact of an interprofessional shared decision-making and goal-setting decision aid for patients with diabetes on decisional conflict--study protocol for a randomized controlled trial. *Trials* 2015 Jun 27;16:286 [FREE Full text] [doi: [10.1186/s13063-015-0797-8](https://doi.org/10.1186/s13063-015-0797-8)] [Medline: [26116444](https://pubmed.ncbi.nlm.nih.gov/26116444/)]
24. Austin Z, Gregory PAM, Martin JC. Negotiation of interprofessional culture shock: The experiences of pharmacists who become physicians. *Journal of Interprofessional Care* 2009 Jul 06;21(1):83-93. [doi: [10.1080/13561820600874817](https://doi.org/10.1080/13561820600874817)] [Medline: [17365376](https://pubmed.ncbi.nlm.nih.gov/17365376/)]
25. Rixon S, Braaf S, Williams A, Liew D, Manias E. Pharmacists' Interprofessional Communication About Medications in Specialty Hospital Settings. *Health Commun* 2015 Oct;30(11):1065-1075. [doi: [10.1080/10410236.2014.919697](https://doi.org/10.1080/10410236.2014.919697)] [Medline: [25317781](https://pubmed.ncbi.nlm.nih.gov/25317781/)]
26. Carayon P, Karsh BT. Agency for Healthcare Research and Quality. Rockville, MD: U.S. Department of Health and Human Services; 2010. Incorporating Health IT Into Workflow Redesign: Request for Information Summary Report URL: <https://healthit.ahrq.gov/sites/default/files/docs/citation/workflowredesignrfi.pdf> [accessed 2018-08-09] [WebCite Cache ID [71XoXpqlK](https://www.webcitation.org/71XoXpqlK)]
27. Singer A, Duarte FR. The effect of electronic medical record system use on communication between pharmacists and prescribers. *BMC Fam Pract* 2015 Oct 28;16:155 [FREE Full text] [doi: [10.1186/s12875-015-0378-7](https://doi.org/10.1186/s12875-015-0378-7)] [Medline: [26507839](https://pubmed.ncbi.nlm.nih.gov/26507839/)]
28. Nugus P, Greenfield D, Travaglia J, Westbrook J, Braithwaite J. How and where clinicians exercise power: Interprofessional relations in health care. *Social Science & Medicine* 2010 Sep;71(5):898-909. [doi: [10.1016/j.socscimed.2010.05.029](https://doi.org/10.1016/j.socscimed.2010.05.029)] [Medline: [20609507](https://pubmed.ncbi.nlm.nih.gov/20609507/)]
29. McDonough RP, Bennett MS. Improving communication skills of pharmacy students through effective precepting. *Am J Pharm Educ* 2006 Jun 15;70(3):58 [FREE Full text] [Medline: [17136179](https://pubmed.ncbi.nlm.nih.gov/17136179/)]
30. Oliver C, Nesbit S, Kelly N. Dissolving Dualisms: How Two Positivists Engaged with Non-Positivist Qualitative Methodology. *International Journal of Qualitative Methods* 2013 Feb;12(1):180-194. [doi: [10.1177/160940691301200106](https://doi.org/10.1177/160940691301200106)] [Medline: [22973420](https://pubmed.ncbi.nlm.nih.gov/22973420/)]
31. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013;13:117 [FREE Full text] [doi: [10.1186/1471-2288-13-117](https://doi.org/10.1186/1471-2288-13-117)] [Medline: [24047204](https://pubmed.ncbi.nlm.nih.gov/24047204/)]
32. Thurmond VA. The Point of Triangulation. *J Nursing Scholarship* 2001 Sep;33(3):253-258. [doi: [10.1111/j.1547-5069.2001.00253.x](https://doi.org/10.1111/j.1547-5069.2001.00253.x)]
33. Hohl CM, Dankoff J, Colacone A, Afilalo M. Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department. *Ann Emerg Med* 2001 Dec;38(6):666-671. [doi: [10.1067/mem.2001.119456](https://doi.org/10.1067/mem.2001.119456)] [Medline: [11719747](https://pubmed.ncbi.nlm.nih.gov/11719747/)]
34. HIMSS Health Information Exchange and Interoperability & Standards Committees. The Business Case for Interoperability and Health Information Exchange: Analysis and Recommendations. Chicago, IL: Healthcare Information and Management Systems Society (HIMSS); 2014. The Business Case for Interoperability and Health Information Exchange: Analysis and Recommendations URL: <https://www.himss.org/business-case-interoperability-and-health-information-exchange> [accessed 2018-08-09] [WebCite Cache ID [71XoqTXHv](https://www.webcitation.org/71XoqTXHv)]
35. Kannampallil TG, Schauer GF, Cohen T, Patel VL. Considering complexity in healthcare systems. *Journal of Biomedical Informatics* 2011 Dec;44(6):943-947. [doi: [10.1016/j.jbi.2011.06.006](https://doi.org/10.1016/j.jbi.2011.06.006)] [Medline: [21763459](https://pubmed.ncbi.nlm.nih.gov/21763459/)]
36. Bardet J, Vo T, Bedouch P, Allenet B. Physicians and community pharmacists collaboration in primary care: A review of specific models. *Research in Social and Administrative Pharmacy* 2015 Sep;11(5):602-622. [doi: [10.1016/j.sapharm.2014.12.003](https://doi.org/10.1016/j.sapharm.2014.12.003)] [Medline: [25640887](https://pubmed.ncbi.nlm.nih.gov/25640887/)]
37. Gregory PAM, Austin Z. Trust in interprofessional collaboration: Perspectives of pharmacists and physicians. *Can Pharm J (Ott)* 2016 Jul;149(4):236-245 [FREE Full text] [doi: [10.1177/1715163516647749](https://doi.org/10.1177/1715163516647749)] [Medline: [27540406](https://pubmed.ncbi.nlm.nih.gov/27540406/)]
38. de Jong CC, Ros WJ, van Leeuwen M, Schrijvers G. How Professionals Share an E-Care Plan for the Elderly in Primary Care: Evaluating the Use of an E-Communication Tool by Different Combinations of Professionals. *J Med Internet Res* 2016 Nov 24;18(11):e304 [FREE Full text] [doi: [10.2196/jmir.6332](https://doi.org/10.2196/jmir.6332)] [Medline: [27884811](https://pubmed.ncbi.nlm.nih.gov/27884811/)]
39. Ancker JS, Wittman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. *J Med Internet Res* 2015 Jun 04;17(6):e137 [FREE Full text] [doi: [10.2196/jmir.4381](https://doi.org/10.2196/jmir.4381)] [Medline: [26043709](https://pubmed.ncbi.nlm.nih.gov/26043709/)]

40. Dogba MJ, Menear M, Stacey D, Brière N, Légaré F. The Evolution of an Interprofessional Shared Decision-Making Research Program: Reflective Case Study of an Emerging Paradigm. *Int J Integr Care* 2016 Jul 19;16(3):4 [FREE Full text] [doi: [10.5334/ijic.2212](https://doi.org/10.5334/ijic.2212)] [Medline: [28435417](https://pubmed.ncbi.nlm.nih.gov/28435417/)]
41. Patel SR, Bakken S, Ruland C. Recent advances in shared decision making for mental health. *Curr Opin Psychiatry* 2008 Nov;21(6):606-612 [FREE Full text] [doi: [10.1097/YCO.0b013e32830eb6b4](https://doi.org/10.1097/YCO.0b013e32830eb6b4)] [Medline: [18852569](https://pubmed.ncbi.nlm.nih.gov/18852569/)]
42. Morant N, Kaminskiy E, Ramon S. Shared decision making for psychiatric medication management: beyond the micro-social. *Health Expect* 2016 Oct;19(5):1002-1014 [FREE Full text] [doi: [10.1111/hex.12392](https://doi.org/10.1111/hex.12392)] [Medline: [26260361](https://pubmed.ncbi.nlm.nih.gov/26260361/)]
43. Elwyn G, Edwards A, Gwyn R, Grol R. Towards a feasible model for shared decision making: focus group study with general practice registrars. *BMJ* 1999 Sep 18;319(7212):753-756. [doi: [10.1136/bmj.319.7212.753](https://doi.org/10.1136/bmj.319.7212.753)]

## Abbreviations

**ADE:** adverse drug effect  
**EHR:** electronic health record  
**EMR:** electronic medical record  
**HCP:** health care professional  
**IP-SDM:** interprofessional shared decision-making  
**SDM:** shared decision-making  
**SHARE:** Secure Health Access Record

*Edited by G Eysenbach; submitted 19.01.18; peer-reviewed by A Hall, H Singh; comments to author 01.03.18; revised version received 23.04.18; accepted 24.06.18; published 25.09.18.*

*Please cite as:*

Mercer K, Burns C, Guirguis L, Chin J, Dogba MJ, Dolovich L, Guénette L, Jenkins L, Légaré F, McKinnon A, McMurray J, Waked K, Grindrod KA

*Physician and Pharmacist Medication Decision-Making in the Time of Electronic Health Records: Mixed-Methods Study*  
*JMIR Hum Factors* 2018;5(3):e24

URL: <http://humanfactors.jmir.org/2018/3/e24/>

doi: [10.2196/humanfactors.9891](https://doi.org/10.2196/humanfactors.9891)

PMID: [30274959](https://pubmed.ncbi.nlm.nih.gov/30274959/)

©Kathryn Mercer, Catherine Burns, Lisa Guirguis, Jessie Chin, Maman Joyce Dogba, Lisa Dolovich, Line Guénette, Laurie Jenkins, France Légaré, Annette McKinnon, Josephine McMurray, Khrystine Waked, Kelly A Grindrod. Originally published in *JMIR Human Factors* (<http://humanfactors.jmir.org>), 25.09.2018. 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 <http://humanfactors.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# A Text Message–Based Intervention Targeting Alcohol Consumption Among University Students: User Satisfaction and Acceptability Study

Ulrika Müssener<sup>1</sup>, PhD; Kristin Thomas<sup>1</sup>, PhD; Catharina Linderöth<sup>1</sup>, PhD; Matti Leijon<sup>1</sup>, PhD; Marcus Bendtsen<sup>1</sup>, PhD

Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

**Corresponding Author:**

Ulrika Müssener, PhD

Department of Medical and Health Sciences

Linköping University

Linköping, Sweden

S-58183

Sweden

Phone: 46 732702426

Fax: 46 13281010

Email: [ulrika.mussener@liu.se](mailto:ulrika.mussener@liu.se)

## Abstract

**Background:** Heavy consumption of alcohol among university students is a global problem, with excessive drinking being the social norm. Students can be a difficult target group to reach, and only a minority seek alcohol-related support. It is important to develop interventions that can reach university students in a way that does not further stretch the resources of the health services. Text messaging (short message service, SMS)–based interventions can enable continuous, real-time, cost-effective, brief support in a real-world setting, but there is a limited amount of evidence for effective interventions on alcohol consumption among young people based on text messaging. To address this, a text messaging–based alcohol consumption intervention, the Amadeus 3 intervention, was developed.

**Objective:** This study explored self-reported changes in drinking habits in an intervention group and a control group. Additionally, user satisfaction among the intervention group and the experience of being allocated to a control group were explored.

**Methods:** Students allocated to the intervention group (n=460) were asked about their drinking habits and offered the opportunity to give their opinion on the structure and content of the intervention. Students in the control group (n=436) were asked about their drinking habits and their experience in being allocated to the control group. Participants received an email containing an electronic link to a short questionnaire. Descriptive analyses of the distribution of the responses to the 12 questions for the intervention group and 5 questions for the control group were performed.

**Results:** The response rate for the user feedback questionnaire of the intervention group was 38% (176/460) and of the control group was 30% (129/436). The variation in the content of the text messages from facts to motivational and practical advice was appreciated by 77% (135/176) participants, and 55% (97/176) found the number of messages per week to be adequate. Overall, 81% (142/176) participants stated that they had read all or nearly all the messages, and 52% (91/176) participants stated that they were drinking less, and increased awareness regarding negative consequences was expressed as the main reason for reduced alcohol consumption. Among the participants in the control group, 40% (52/129) stated that it did not matter that they had to wait for access to the intervention. Regarding actions taken while waiting for access, 48% (62/129) participants claimed that they continued to drink as before, whereas 35% (45/129) tried to reduce their consumption without any support.

**Conclusions:** Although the main randomized controlled trial was not able to detect a statistically significant effect of the intervention, most participants in this qualitative follow-up study stated that participation in the study helped them reflect upon their consumption, leading to altered drinking habits and reduced alcohol consumption.

**Trial Registration:** International Standard Randomized Controlled Trial Number ISRCTN95054707; <http://www.isrctn.com/ISRCTN95054707> (Archived by WebCite at <http://www.webcitation.org/705putNZT>)

(*JMIR Hum Factors* 2018;5(3):e23) doi:[10.2196/humanfactors.9641](https://doi.org/10.2196/humanfactors.9641)

**KEYWORDS**

students; text messages; mobile phones; SMS; alcohol consumption intervention

## Introduction

A large proportion of the global burden of disease is due to excessive alcohol consumption. Alcohol-related deaths increased by 30%, or approximately 5 million, between 1990 and 2010 [1]. Despite these health risks, heavy alcohol consumption among university students remains a global problem, with excessive drinking being the social norm [2,3]. In addition, research shows that students can be a difficult target group to reach, and only a minority seek alcohol-related support. Typically, local on-site student health services are commissioned to offer preventative services as well as advice and support to students who wish to reduce or discontinue drinking. However, these student health services must do so with limited resources [4]. Thus, it is important to develop interventions that can reach university students in a way that does not further stretch the resources of the health services.

Research has shown that interventions delivered by text messaging, also known as short message service (SMS), is a cost-effective method to support behavioral change [5] such as weight loss, smoking cessation, and diabetes management [6,7]. For instance, a 12-week text messaging-based intervention targeting heavy drinking among young adults was found to influence the number of days of heavy drinking and the number of drinks per drinking day [8].

Moreover, positive evidence regarding usability and user experience of text messaging-based interventions has been observed. For instance, text messages have been shown to be highly accessible to users in the sense that messages are likely to be read within minutes of being received, and interventions have been shown to be user-friendly as reading text messages requires limited time and effort [9-11]. Thus, text messaging-based interventions can enable continuous, real-time, brief support in a real-world setting [9,12,13].

This study builds on an earlier randomized controlled trial [14,15] that aimed to show the effect of using a text messaging-based alcohol consumption intervention among university students. This study has three aims:

1. To explore self-reported changes in drinking habits in an intervention group and a control group;
2. To explore user satisfaction among the intervention group given access to the novel intervention;
3. To explore the experience of being allocated to the control group.

## Methods

Ethical approval for this randomized controlled trial (RCT, ISRCTN95054707) was given by the Regional Ethical Committee in Linköping, Sweden (dnr 2016/134-31).

### Short Description of the Amadeus 3 Intervention

The Amadeus 3 intervention was developed using formative methods, including focus groups with students, an expert panel

with students and professionals, and behavioral change technique analysis. The development of the program has been previously described [16]. The intervention included facts about the negative consequences of alcohol, tips on behavioral change strategies, and activities such as saying no to alcohol.

The intervention consisted of a 6-week program with a total of 62 messages. At the start, users were asked to set a goal of how much they would like to reduce their drinking. The first 4 weeks of the program had a higher frequency of 9 messages each week, followed by 7 messages in week 5, and 5 messages in week 6, all together 48 messages. Messages were sent at various times around midday, late afternoon, or early evening. Of the 62 messages, 48 were unique and 14 messages were repeated. Two messages were repeated at the start of each week, as students were asked to report via a text the number of drinks they had consumed the previous week. Following their response, they received a second text including feedback on their performance in relation to the goal they set at the start of the intervention. These paired messages were repeated every Sunday. The content of the unique messages was primarily based on information or behavioral practice. Information-based messages typically included facts about alcohol and health, consequences of excessive drinking, or tips on behavioral change strategies. Behavioral practice-based messages included asking students to reflect on or practice behavioral change; for instance, students were asked to reflect on triggers for excessive drinking or practicing saying no to drinking on a night out [16].

The messages were sent from a GSM modem and administered from a technical platform that was developed and owned by one of the authors (MB). Sending of all text messages was fully automated using this platform.

### Study Population and Recruitment

University students participating in the main Amadeus 3 study were invited to give feedback after completing the 6-week intervention and participating in the formal follow-up of the RCT [15]. Participants were recruited from 13 colleges and universities in Sweden. A total of 460 participants were allocated to the intervention group and 436 to the control group, which was offered treatment as usual (eg, other support provided at the universities such as advice and support from student health care). In the main study, follow-up data on the primary outcome were collected from 423 participants (92%) of the intervention group and 392 participants (90%) of the control group. Two reminders to complete the follow-up questionnaire were sent by email at 1 and 2 weeks after the initial request. Nonresponders were then sent text message reminders every other day for 6 days (a total of 3 text messages), and finally were contacted by phone (with a maximum of 10 calls).

After the follow-up procedure of the RCT, a second questionnaire was sent to both groups. The intervention group was asked about drinking habits and offered the opportunity to give their opinion on the structure and content of the intervention. The control group was asked about drinking habits

and their experience in being allocated to the control group. The questionnaire was sent by email with 2 weekly reminders.

### Questionnaire

The intervention group was asked 12 questions. Each question had 2–7 fixed-response options with an optional free-text comment field, except for question 2 for which only a free-text comment was offered. Free-text comments gave participants an opportunity to describe other factors of importance not covered by the fixed-response options.

Initially, drinking habits were explored by 2 questions: (1) change in drinking habits during participation in the program (response options: I drink more, I drink less, I drink the same amount as before, I stopped drinking, I don't know) and (2) possible reasons for having stopped drinking or drinking less if applicable (only free-text comment).

Experiences with the structure of the intervention were explored by 5 questions: (1) defining a goal for weekly consumption at the beginning of the program (response options: very good/good/neither good nor bad/bad/very bad/don't know), (2) the mix of motivating, supporting, and factual content (response options: very good/good/neither good nor bad/bad/very bad/I don't know), (3) how the participants experienced the duration of the intervention (response options: far too long /somewhat too long/just right/somewhat too short/too short/don't know), (4) how the participants experienced the number of messages per week (response options: far too many/somewhat too many/just right/somewhat too few/too few/don't know), and (5) how long after receiving the messages did the participants actually read them (response options: immediately/within 1 hour/within a couple of hours/same day/next day).

Experience with the content of the intervention was explored by 5 questions: (1) the content of the messages (response options: very good/good/neither good nor bad/bad/don't know), (2) the proportion of the messages that the participant perceived to be useful (response options: all/nearly all/about half/some/nearly none/none/don't know), (3) the proportion of all messages that were read (response options: all/nearly all/about half/some/nearly none/none/don't know), (4) whether the participant would recommend the intervention to a friend who should reduce alcohol consumption (response options: yes/unsure/no/don't know), and (5) whether the participant had used any additional support during the intervention (response options: no/yes).

Participants from the control group were asked 5 questions. Each question had 4 or 5 fixed-response options and offered an optional free-text comment field, except for question 2 for which only a free-text comment was provided.

Drinking habits of the control group were explored by the same 2 questions as for the intervention group: (1) change in drinking habits since the beginning of the trial (response options: I drink more, I drink less, I drink the same amount as before, I stopped

drinking, I don't know) and (2) possible reasons for having stopped drinking or drinking less if applicable (only free-text comment).

Experience with and actions taken from being randomized to the control group were explored by 2 questions: (1) experience of having to wait for support from the program (response options: disappointed because I expected to get support immediately/ok because I had time to reflect upon my alcohol habits/didn't matter/don't know) and (2) actions taken while waiting for support from the program (response options: I used other support [type of support was to be specified]/I decided to reduce my consumption until I got support from the program/I tried to reduce my consumption without support/I continued to drink as before/don't know). The final question explored whether control group participants felt that the information regarding the study design was sufficient when signing up (response options: yes, very good/yes, ok/no I lacked information [type of information was to be specified]/don't know).

### Data Analysis

Descriptive analyses of the distribution of the responses to the 12 questions for the intervention group and 5 questions for the control group were performed. In the first step of the analyses, all free-text comments to each question were read by 2 authors (UM and CL). In the second step, UM chose a variety of the most crucial free-text comments for each question. In the third step of the analysis, UM presented the chosen comments to the other authors and, after discussion, the comments that captured the main content of the specific question regarding the aim of the study were chosen. The free-text comments were used to underscore and illustrate the pattern of responses to the fixed-response options. The number after each comment represented the individual code that was assigned to each of the respondents. ".../" showed that part of the free-text comment had been omitted.

## Results

### Overview

Baseline data were used to assess differences between responders and nonresponders. Variables included sex, age, marital status, total number of standard drinks consumed per week, number of episodes of heavy drinking, highest estimated blood alcohol concentration, and the number of negative consequences experienced. The response rate for the intervention group was 38.2% (176/460). As can be seen in [Table 1](#), the baseline characteristics of responders were similar to those of the nonresponders except for sex; the data indicated that females were over-represented in the responders group. The response rate for the control group was 29.6% (129/436). No significant differences were found among participants and nonparticipants with respect to baseline characteristics ([Table 2](#)).



**Table 1.** Baseline characteristics of the responders and nonresponders in the intervention group follow-up.

Variable	Nonresponders (n=284)	Responders (n=176)	P value
Sex, n (% female)	157 (55.3)	108 (61.4)	.001 <sup>a</sup>
Age (years), median (IQR) <sup>b</sup>	24 (5.5)	24 (3.5)	NS <sup>c</sup>
Marital status, n (% single)	182 (64.1)	106 (60.2)	NS
Total weekly alcohol consumption (in standard drinks), median (IQR)	12 (10)	12 (8.25)	NS
<b>Heavy episodic drinking, n (%)</b>			NS
Two or three times a month	79 (27.8)	32 (18.2)	
Once or twice a week	149 (52.5)	92 (52.3)	
Three times or more a week	56 (19.7)	52 (29.5)	
eBAC <sup>d</sup> , median (IQR)	1.37 (1.11)	1.32 (1.12)	NS
Number of negative consequences, median (IQR)	3 (3)	3 (3)	NS

<sup>a</sup>Pearson's chi-square test indicated that the distributions in the 2 groups differed significantly ( $\chi^2_1=10.652, P=.001$ ).

<sup>b</sup>IQR: interquartile range.

<sup>c</sup>NS: not significant.

<sup>d</sup>eBAC: estimated blood alcohol concentration.

**Table 2.** Baseline characteristics of the responders and nonresponders in the control group follow-up.

Variables	Nonresponders (n=307)	Responders (n=129)	P value
Sex, n (% female)	172 (56.0)	72 (55.8)	NS <sup>a</sup>
Age (years), median (IQR) <sup>b</sup>	24 (6)	23 (5)	NS
Marital status, n (% single)	185 (60.3)	71 (55.0)	NS
Total weekly alcohol consumption (in standard drinks), median (IQR)	12 (10)	11 (8)	NS
<b>Heavy episodic drinking, n (%)</b>			NS
Two or three times a month	87 (28.3)	43 (33.3)	
Once or twice a week	150 (48.9)	68 (52.7)	
Three times or more a week	70 (22.8)	18 (14.0)	
eBAC <sup>c</sup> , median (IQR)	1.48 (1.22)	1.33 (1.35)	NS
Number of negative consequences, median (IQR)	3 (3)	2 (3)	NS

<sup>a</sup>NS: not significant.

<sup>b</sup>IQR: interquartile range.

<sup>c</sup>eBAC: estimated blood alcohol concentration.

In the intervention group, 70.4% (124/176) of the participants provided at least one comment to the 12 questions and the other 29.5% (52/176) did not offer any additional comments. In the intervention group, 54.5% (96/176) participants provided comments on possible reasons for having stopped drinking or drinking less, 15.9% (28/176) on the question about change in drinking habits during the program, and 2.3% (4/176) on the question on the time between receiving and reading the messages. On average, approximately 20 comments were provided for each question.

In the control group, 50.3% (65/129) of the participants provided comments to the 5 questions, and the other 49.6% (64/129) did not offer any additional comments. As in the intervention group, most comments were on possible reasons for having stopped drinking or drinking less (42.6%, 55/129) and the question on

changes in drinking habits during the program (7.0%, 9/129). The fewest number of comments was provided for the question on whether the participants found that the information regarding the study design was sufficient when signing up 4.6% (6/129). On average, around 17 comments were provided for each question.

We report the responses to the relevant questions and include citations from the free-text comments for each heading.

### Changes in Drinking Habits and Reasons for Reduced Consumption

#### Intervention Group

In the intervention group, 34.6% (61/176) of participants reported that they consumed the same amount of alcohol as that

before the intervention, 51.7% (91/176) stated that they were drinking less, 4.5% (8/176) stated that they had stopped drinking altogether, 4.0% (7/176) stated that they were drinking more than before, and 5.1% (9/176) answered that they did not know. In the free-text comments, some participants expressed that participation in the study helped them reflect on their alcohol consumption, leading to changed drinking habits.

*I don't want to stop drinking at a party, but I reflect more on my consumption since I started participating in the study. So participation has probably affected me positively [2540432]*

*I reflect about my consumption more. I have fewer binges, but I still drink far too much at times and get memory loss. [254551]*

*Not the greatest change, but for example, I don't drink at home before the party any more. [2545551]*

Increased awareness of negative consequences was expressed as the main reason for reduced alcohol consumption among participants in the intervention group who reduced or stopped drinking during the intervention. In their comments, many described having experienced consequences regarding economics, health, relationships, and exam results.

*I'm more aware of the negative consequences that come with drinking and for me that means that I now want to start living a healthier life. [2540768]*

*I think more about what alcohol does to my health, the economy, relationships, work, etc [2540888]*

*I reflect more about my drinking nowadays. And I think more about how much I drank or how much I planned to drink and how that affects me. Thus, I sometimes completely refrain from alcohol or choose to drink less when something is to be celebrated. I received a lot of good advice from the study. [2540874]*

Increased motivation, loss of control when drinking, and emotional consequences of alcohol consumption were also mentioned in the free-text comments.

*Motivation. .../I say and do things that I'm ashamed of the day after drinking. I get anxiety and I get very sick the day after. [2568730]*

*Because I don't like to get these memory shutters. What made me think was when you had to count the units of alcohol and I saw how much I drank. [2537333]*

Some participants also described their reduction in alcohol consumption because of life changes, such as getting pregnant, moving to another city, entering working life, as well as new social networks and relationships.

*Change of habits because of a new job. [2540674]*

*Changed living situation, I don't live in a student city any longer. [2541284]*

### Control Group

Of the participants in the control group, 41.1% (53/129) reported that they drank the same amount as that before the trial, 32.5%

(42/129) stated that they were drinking less, 1.5% (2/129) stated that they had stopped drinking alcohol, 11.6% (15/129) stated that they drank more than before, and 13.2% (17/129) answered that they did not know. The reasons for reduced consumption seem to be similar to those of the intervention group, such as awareness of their drinking habits, which in turn affected changes in lifestyle.

*I don't hang out with the same people anymore /.../ I don't have the same drinking habits that students in Norrkoping have [2154339]*

*I feel better. I have a healthier relationship with alcohol. I live together with my partner now. I focus more on my studies, training, and health. [2539339]*

### Satisfaction With the Structure of the Intervention

The intervention began with a request to all participants in the intervention group to define a goal for their drinking habits. Of the participants, 77.8% (137/176) agreed that having to define a goal was good or very good at the beginning of the intervention.

*I think it was good because you always thought about that goal. Once you drank, you thought about exactly how much you drank, which caused you to drink less. [2540464]*

*Well, it made me feel guilty when I answered the text messages and wrote how much I had been drinking this week. [2568862]*

On the other hand, 2.3% (4/176) of the participants reported that setting a goal was bad, and one participant commented that it created an expectation from the participant and therefore had the opposite effect.

*I felt concerned about the demands or expectations, and it had the opposite effect [2541027]*

The variation in the content of the text messages from facts to motivational and practical advice was appreciated by 76.7% of the participants (135/176), particularly among those who reduced or quit drinking. Some participants emphasized the need for different content because people have different needs; some wanted more facts, and 34.6% (61/176) said the variation was bad or very bad.

*A good variety in order to cover the different areas that may cause problems. [2540491]*

*People do handle things differently - then variety is good. [2544151]*

*I liked when there were facts, eg, how alcohol affect sleep. Even more facts like that would be good! [2546086]*

The duration of the 6-week intervention was perceived to be adequate by 55.1% of the participants (97/176).

*During the program, I thought it was an unnecessarily long number of weeks and not very fulfilling. It was only afterward that I began to think that it was quite worthwhile. I think the length of it was the most rewarding aspect for me, to actively reflect on alcohol for a relatively long time. [2543442]*

Of the participants, 17.0% (30/176) stated that it was too short, and 19.9% (35/176) stated that it was too long. One participant suggested sending less than one text message per day.

*No problem with 6 weeks, but maybe not with text messages every day. It felt almost tedious. For me, the problem arose at the weekends when there was a party, and that is the time when you might need the text messages as a reminder. [2541735]*

Of the participants, 54.5% (96/176) found the number of messages per week to be about right; 34.1% (60/176) thought there were too many messages. One participant said that they looked forward to the daily text message, but that there was a lot of information to digest.

*Every day I looked forward to the text messages, but on the other hand it was very extensive /.../there was a lot of information to take in [2540677]*

On average, 19.9% (35/176) of participants read the text messages immediately when they received them, 53.4% (94/176) read the messages within 1 hour, 15.3% (27/176) read the messages within a couple of hours, and 2.3% (4/176) of participants read the text messages the next day.

### Satisfaction With the Content of the Intervention

Regarding the content of the text messages, 64.2% of participants (113/176) in the intervention group found the content good or very good. Among those who were in favor of the content, some emphasized that the messages changed how they thought about alcohol consumption, and they were reminded about why they wanted to reduce or stop drinking. Others stated that the messages made them reflect on their drinking in a more conscious way.

*Interesting information. I had expected horror propaganda, but many of the issues were about putting your alcohol problems in perspective and reflecting on your habits [2540667]*

Of the participants, 6.2% (11/176) found the content bad or very bad. Some participants who were still drinking as before perceived the messages as irritating and impersonal. One participant described the messages as a bad joke.

*Aggravating. It was like a bad joke, obvious and impersonal. But the program gave us good laughs at our party anyway. I thought it was so bad that I ended in advance. [2544204]*

A total of 27.8% (49/176) did not find the content either good or bad. Some emphasized that the actual content of the messages was not very important; rather, it was more important to be reminded and encouraged to think about one's alcohol consumption.

*I did not feel like it was the content itself that mattered, many of the text messages repeated things I already know. For me what made the difference, however, was being reminded to be aware of my plans for drinking - which I did by reading the text messages. What was in the messages did not matter. And, it felt like someone was supervising a little when*

*the messages came, which also made me less motivated to drink. [2539338]*

Some thought that the messages were interesting and relevant initially but then became repetitive, and they suggested shorter messages less often and that the messages should be sent out earlier in the evenings, before they started to drink. More supporting and motivating messages before weekends were appreciated.

*You had the chance to analyze your decision to drink, so sometimes you chose not to drink alcohol because you were affected by the message. [2543929]*

The proportion of messages that the participants perceived to be useful differed. Only 35.8% of participants (63/176) thought that all or nearly all messages were useful for their situation. Some experienced that the messages gave a feeling that somebody cared.

*I was thankful for the messages in any case. It felt like someone cared, although virtual. [2540886]*

Among those who were still drinking the same amount as before the intervention and who estimated that none of the messages were useful, one participant said that the messages had the opposite effect: a desire to drink.

*Seemed that the messages were far too goading and made me think of alcohol more and to drink more, the opposite of their purpose [2541098]*

In all, 80.7% of participants (142/176) stated that they had read all or nearly all the messages and only 7.9% of participants (14/176) reported that they only read a few or none of the messages. Furthermore, 48.3% of the participants (85/176) said that they would recommend the intervention to a friend, 39.2% (69/176) said that they were unsure or did not know if they would recommend the intervention, and 12.5% (22/176) would not recommend it.

*Uncertain. If one really needs help, more action is required. But it can be a good way to reflect on ones' drinking as well as being reminded. [2540432]*

*Yes, but only if it's a person who has previously thought about reducing his or her alcohol consumption. For a person who had not reflected on it earlier, the program would probably not be so useful. [2540768]*

Of the participants, 95.4% (168/176) had not used additional support during the intervention. Seven free-text comments mentioned the use of the following additional support: reading a book regarding the power of habits, face-to-face encounters with professionals, dialog with relatives, medication, and therapy.

### Experiences of Being Randomized Into the Control Group

Among the participants in the control group, 40.3% (52/129) expressed that it did not matter that they had to wait for access to the intervention, and 20.9% (27/129) stated that it was fine to wait because it gave them time to reflect on their consumption, whereas 27.9% of participants (36/129) expressed disappointment with having to wait for support.

*Of course, I wanted to get the support as quickly as possible, but there were no big problems having to wait. [546113]*

*However, having to wait seemed a bit frustrating at first. [2544776]*

Regarding actions taken while waiting for access, 48.1% of the participants (62/129) claimed that they continued to drink as before, and 34.9% (45/129) tried to reduce their consumption without any support; 3.1% of participants (4/129) reported that they decided to wait until they were given access to the intervention, and the same number of participants reported that they used other aids, such as medication and support from the alcohol-dependence units.

The final question explored whether the participants in the control group found that the information regarding the study design was sufficient when signing up. Of the participants, 67.4% (87/129) stated that the information was good or very good, 27.1% (35/129) answered that they did not know, and 5.4% (7/129) stated that they did not think that the information was sufficient.

*You did not really know what you were getting yourself into; the study was quite unclear and it was hard to know when the 3 months had passed or were still ongoing [2540443]*

## Discussion

### Principal Findings

The main findings of this study are that most of the participants in the intervention group state that participation in the study helps them to reflect on their consumption, leading to altered drinking habits and reduced alcohol consumption. Most of the participants appreciate the variation in the content of the text messages from facts to motivational and practical advice. The results also shine light on the experience of being allocated to a control group and that it does not matter that they must wait for access to the intervention.

Despite the low response rate of 34%, participants in this user evaluation study provided valuable information regarding alcohol consumption and changes in drinking habits, as well as user satisfaction and the experience of being allocated to a control group that can be used in further work regarding alcohol-related support. A note of caution, however, should be made. There is an over-representation of females who responded in the intervention group (Table 1). Apart from this, participants are broadly representative of the study population (participants in the RCT).

Many individuals find the support helpful, yet in terms of being effective toward a general nontreatment-seeking population, there is more work to be done. The reasons given for reduced consumption seem to be the same in both the control and intervention groups: increased awareness regarding negative consequences for health, economics, relationships, study results, and work, as well as changes in civil status. However, it is not possible to say why the reasons are the same in both groups. New interventions need to focus on the factors that both groups estimate to be important when it comes to changes in lifestyle.

Previous research shows that despite the promising potential of text messaging-based interventions, it is difficult to tell how effectiveness may be optimized through the content and structure choices [11,17]. This study sheds light on some of these questions because the results show that the overall structure and content of the text messaging-based intervention are well received by most participants, regardless of whether they reduce their drinking. Most agree that the variation is valuable, particularly among those who reduced or stopped drinking. One possible conclusion, also noted in previous research using a text messaging-based intervention among students who smoked, is that text messaging-based cessation interventions are more suitable for those who are motivated to use these types of programs, and those who are not fully motivated or determined to change their lifestyle habits may find other types of support more suitable, for example face-to-face meetings with professionals [18].

Participants in the intervention group who appreciated the content stressed that the messages changed their thinking about alcohol consumption. Among participants who were less appreciative of the content, some emphasized that the content itself is of less importance. Instead, most of the gain is in being reminded and encouraged to reflect on one's consumption. It is unclear if it is the content of the messages or the frequent reminders and reinforcement of having committed oneself to reduce one's drinking that matters. Similar results were shown in a previous study using a text messaging-based intervention to stop smoking among young people [19], and mechanisms of the effect of this type of intervention remain to be identified and studied further. The remarks that the same messages are perceived as irritating, impersonal, and repetitive by some and useful by others reflect the limitations imposed by untailed interventions and highlight the difficulty in developing an intervention that fits all, an issue that has been discussed extensively [20]. The variation in content is appreciated by most participants, but the proportion of messages that the participants perceived to be useful differs. The feeling of being cared for is mentioned as important among those who think that the messages are valuable to their situation.

Additional support is used by few, and because the intervention should be a complement to other support provided at the universities, the results are not affected in any direction. Previous research shows that only a minority of students seek advice and support from student health care, and our results emphasized the need to further develop new means of reaching students who drink excessively [21].

Being randomized into a control group implies having to wait for support for approximately 3 months, but 40.3% of the participants (52/129) expressed that delayed support did not matter. Some used the waiting time to reflect on their consumption. Concern regarding the ethics of assigning participants actively ready for change to control groups has been raised [22] and asking participants in control groups to wait to seek treatment may lessen their natural help-seeking behaviors [23]. However, participants in this study were free to seek other treatments, but only a few chose to do so; indeed, half of the participants continued to drink as before.

The strengths of this study are that participants were recruited from 14 colleges and universities in Sweden and that there were many free-text comments to most of the questions. The intervention is fully automated and did not require the user to remember to log in to a web portal or similar website throughout the intervention.

### Limitations

Limitations of the study include the low response rate and the relatively short questionnaires used to explore the views of the participants. The duration of the intervention was adequate for only about half of the participants, indicating that the optimal duration of the intervention is still to be established.

Several steps have been taken to ensure the validity of the results. Two authors read the free-text comments independently many times. The first author selected a variety of the most crucial free-text comments, and then the chosen free-text comments were presented and discussed with the other authors, and comments that captured the main content of the specific

question about the aim of the study were chosen. Free-text comments not agreed on by all authors were excluded.

### Conclusions

Reflecting on alcohol consumption may help young people change their drinking habits and reduce their alcohol consumption. Variation in the content of the intervention from facts to motivational and practical advice seems to be satisfactory, but the optimal duration of the intervention, as well as the number of messages per week, is still to be established. Further work is needed to determine what aspects matter to support students who wish to reduce or quit drinking. To obtain such knowledge, students' experiences are probably highly significant, especially in the context of improving understanding of the mechanisms behind a successful text messaging-based intervention. Deeper knowledge is needed about whether it is the content itself that is important or if the gain is in being reminded frequently and encouraged to reflect on one's consumption.

---

### Acknowledgments

The study was funded by the Public Health Agency of Sweden.

---

### Authors' Contributions

UM, KT, and MB designed the study and developed the intervention. CL took a leading role in recruiting the participants. MB did all the programming and supervised the data collection and all technical aspects of the delivery of the intervention. UM wrote the first draft of the manuscript and revised it with input from first ML and then all authors.

---

### Conflicts of Interest

None declared.

---

### References

1. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: [10.1016/S0140-6736\(12\)61766-8](https://doi.org/10.1016/S0140-6736(12)61766-8)] [Medline: [23245609](https://pubmed.ncbi.nlm.nih.gov/23245609/)]
2. Andersson C, Johnsson KO, Berglund M, Ojehagen A. Intervention for hazardous alcohol use and high level of stress in university freshmen: a comparison between an intervention and a control university. *Brain Res* 2009 Dec 11;1305 Suppl:S61-S71. [doi: [10.1016/j.brainres.2009.08.030](https://doi.org/10.1016/j.brainres.2009.08.030)] [Medline: [19699184](https://pubmed.ncbi.nlm.nih.gov/19699184/)]
3. Wicki M, Kuntsche E, Gmel G. Drinking at European universities? A review of students' alcohol use. *Addict Behav* 2010 Nov;35(11):913-924. [doi: [10.1016/j.addbeh.2010.06.015](https://doi.org/10.1016/j.addbeh.2010.06.015)] [Medline: [20624671](https://pubmed.ncbi.nlm.nih.gov/20624671/)]
4. Statistics Sweden. Population Census 2014 URL: <http://www.scb.se/sv/Hitta-statistik/Statistik-efter-amne/Befolkning/> [accessed 2018-05-29] [WebCite Cache ID 6zm3v8Vy6]
5. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. *Prev Sci* 2015 Feb;16(2):181-188. [doi: [10.1007/s11121-014-0498-7](https://doi.org/10.1007/s11121-014-0498-7)] [Medline: [24930386](https://pubmed.ncbi.nlm.nih.gov/24930386/)]
6. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. *Epidemiol Rev* 2010;32:56-69 [FREE Full text] [doi: [10.1093/epirev/mxq004](https://doi.org/10.1093/epirev/mxq004)] [Medline: [20354039](https://pubmed.ncbi.nlm.nih.gov/20354039/)]
7. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. *Am J Prev Med* 2009 Feb;36(2):165-173. [doi: [10.1016/j.amepre.2008.09.040](https://doi.org/10.1016/j.amepre.2008.09.040)] [Medline: [19135907](https://pubmed.ncbi.nlm.nih.gov/19135907/)]
8. Suffoletto B, Callaway C, Kristan J, Kraemer K, Clark DB. Text-message-based drinking assessments and brief interventions for young adults discharged from the emergency department. *Alcohol Clin Exp Res* 2012 Mar;36(3):552-560. [doi: [10.1111/j.1530-0277.2011.01646.x](https://doi.org/10.1111/j.1530-0277.2011.01646.x)] [Medline: [22168137](https://pubmed.ncbi.nlm.nih.gov/22168137/)]
9. Douglas N, Free C. 'Someone battling in my corner': experiences of smoking-cessation support via text message. *Br J Gen Pract* 2013 Nov;63(616):e768-e776 [FREE Full text] [doi: [10.3399/bjgp13X674459](https://doi.org/10.3399/bjgp13X674459)] [Medline: [24267860](https://pubmed.ncbi.nlm.nih.gov/24267860/)]

10. Bendtsen M, Bendtsen P. Feasibility of a Fully Automated Multiple Session Alcohol Intervention to University Students, Using Different Modes of Electronic Delivery—The TOPHAT 1 Study. *JSEA* 2013;06(10):14-26. [doi: [10.4236/jsea.2013.610A003](https://doi.org/10.4236/jsea.2013.610A003)]
11. Bendtsen M, Bendtsen P. Feasibility and user perception of a fully automated push-based multiple-session alcohol intervention for university students: randomized controlled trial. *JMIR Mhealth Uhealth* 2014;2(2):e30 [FREE Full text] [doi: [10.2196/mhealth.3233](https://doi.org/10.2196/mhealth.3233)] [Medline: [25098296](https://pubmed.ncbi.nlm.nih.gov/25098296/)]
12. Trull TJ, Ebner-Priemer U. Ambulatory assessment. *Annu Rev Clin Psychol* 2013;9:151-176 [FREE Full text] [doi: [10.1146/annurev-clinpsy-050212-185510](https://doi.org/10.1146/annurev-clinpsy-050212-185510)] [Medline: [23157450](https://pubmed.ncbi.nlm.nih.gov/23157450/)]
13. Beckjord E, Shiffman S. Background for Real-Time Monitoring and Intervention Related to Alcohol Use. *Alcohol Res* 2014;36(1):9-18 [FREE Full text] [Medline: [26258996](https://pubmed.ncbi.nlm.nih.gov/26258996/)]
14. Thomas K, Bendtsen M, Linderoth C, Karlsson N, Bendtsen P, Müssener U. Short message service (SMS)-based intervention targeting alcohol consumption among university students: study protocol of a randomized controlled trial. *Trials* 2017 Apr 04;18(1):156 [FREE Full text] [doi: [10.1186/s13063-017-1898-3](https://doi.org/10.1186/s13063-017-1898-3)] [Medline: [28372563](https://pubmed.ncbi.nlm.nih.gov/28372563/)]
15. Thomas K, Müssener U, Linderoth C, Karlsson N, Bendtsen P, Bendtsen M. Effectiveness of a Text Messaging-Based Intervention Targeting Alcohol Consumption Among University Students: Randomized Controlled Trial. *JMIR Mhealth Uhealth* 2018 Jun 25;6(6):e146 [FREE Full text] [doi: [10.2196/mhealth.9642](https://doi.org/10.2196/mhealth.9642)] [Medline: [29941417](https://pubmed.ncbi.nlm.nih.gov/29941417/)]
16. Thomas K, Linderoth C, Bendtsen M, Bendtsen P, Müssener U. Text Message-Based Intervention Targeting Alcohol Consumption Among University Students: Findings From a Formative Development Study. *JMIR Mhealth Uhealth* 2016 Oct 20;4(4):e119 [FREE Full text] [doi: [10.2196/mhealth.5863](https://doi.org/10.2196/mhealth.5863)] [Medline: [27765732](https://pubmed.ncbi.nlm.nih.gov/27765732/)]
17. Berman AH, Gajeci M, Sinadinovic K, Andersson C. Mobile Interventions Targeting Risky Drinking Among University Students: A Review. *Curr Addict Rep* 2016;3:166-174 [FREE Full text] [doi: [10.1007/s40429-016-0099-6](https://doi.org/10.1007/s40429-016-0099-6)] [Medline: [27226948](https://pubmed.ncbi.nlm.nih.gov/27226948/)]
18. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of Short Message Service Text-Based Smoking Cessation Intervention Among University Students: A Randomized Clinical Trial. *JAMA Intern Med* 2016 Mar;176(3):321-328. [doi: [10.1001/jamainternmed.2015.8260](https://doi.org/10.1001/jamainternmed.2015.8260)] [Medline: [26903176](https://pubmed.ncbi.nlm.nih.gov/26903176/)]
19. Müssener U, Bendtsen M, McCambridge J, Bendtsen P. User satisfaction with the structure and content of the NEXit intervention, a text messaging-based smoking cessation programme. *BMC Public Health* 2016 Nov 22;16(1):1179 [FREE Full text] [doi: [10.1186/s12889-016-3848-5](https://doi.org/10.1186/s12889-016-3848-5)] [Medline: [27876031](https://pubmed.ncbi.nlm.nih.gov/27876031/)]
20. Marsch L, Lord S, Dallert J. Behavioral Healthcare and Technology: Using Science-Based Innovations to Transform Practice. In: Oxford University Press 2014.
21. Cellucci T, Krogh J, Vik P. Help seeking for alcohol problems in a college population. *J Gen Psychol* 2006 Oct;133(4):421-433. [doi: [10.3200/GENP.133.4.421-433](https://doi.org/10.3200/GENP.133.4.421-433)] [Medline: [17128960](https://pubmed.ncbi.nlm.nih.gov/17128960/)]
22. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Med Res Methodol* 2013 Dec 06;13:150 [FREE Full text] [doi: [10.1186/1471-2288-13-150](https://doi.org/10.1186/1471-2288-13-150)] [Medline: [24314204](https://pubmed.ncbi.nlm.nih.gov/24314204/)]
23. Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychother Psychosom* 2009;78(5):275-284. [doi: [10.1159/000228248](https://doi.org/10.1159/000228248)] [Medline: [19602916](https://pubmed.ncbi.nlm.nih.gov/19602916/)]

## Abbreviations

**eBAC:** estimated blood alcohol concentration

**IQR:** interquartile range

**NS:** not significant

**RCT:** randomized clinical trial

**SMS:** short message service

*Edited by G Eysenbach; submitted 12.12.17; peer-reviewed by B Suffoletto, B Heckman; comments to author 18.03.18; revised version received 03.04.18; accepted 10.05.18; published 10.07.18.*

*Please cite as:*

Müssener U, Thomas K, Linderoth C, Leijon M, Bendtsen M

*A Text Message-Based Intervention Targeting Alcohol Consumption Among University Students: User Satisfaction and Acceptability Study*

*JMIR Hum Factors* 2018;5(3):e23

URL: <http://humanfactors.jmir.org/2018/3/e23/>

doi: [10.2196/humanfactors.9641](https://doi.org/10.2196/humanfactors.9641)

PMID: [29991469](https://pubmed.ncbi.nlm.nih.gov/29991469/)

©Ulrika Müssener, Kristin Thomas, Catharina Linderöth, Matti Leijon, Marcus Bendtsen. Originally published in JMIR Human Factors (<http://humanfactors.jmir.org>), 10.07.2018. 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 <http://humanfactors.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# The D2Refine Platform for the Standardization of Clinical Research Study Data Dictionaries: Usability Study

Deepak Kumar Sharma<sup>1</sup>, MS, MCompSc; Kevin Jerrold Peterson<sup>1</sup>, MS; Na Hong<sup>1</sup>, PhD; Guoqian Jiang<sup>1</sup>, MD, PhD

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

**Corresponding Author:**

Guoqian Jiang, MD, PhD

Department of Health Sciences Research

Mayo Clinic

200 First Street SW

Rochester, MN,

United States

Phone: 1 507 284 2511

Email: [jiang.guoqian@mayo.edu](mailto:jiang.guoqian@mayo.edu)

## Abstract

**Background:** D2Refine provides a Web-based environment to create clinical research study data dictionaries and enables standardization and harmonization of its variable definitions with controlled terminology resources.

**Objective:** To assess the usability of the functions D2Refine offers, a usability study was designed and executed.

**Methods:** We employed the TURF (task, user, representation, and function) Usability Framework of electronic health record usability to design, configure, and execute the usability study and performed quantitative analyses. D2Refine was compared for its usability metrics against two other comparable solutions, OntoMaton and RightField, which have very similar functionalities for creating, managing, and standardizing data dictionaries. We first conducted the function analysis by conducting one-on-one interviews armed with questionnaires to catalog expected functionality. The enrolled participants carried out the steps for selected tasks to accomplish specific goals and their feedback was captured to conduct the task analysis.

**Results:** We enrolled a group (n=27) of study developers, managers, and software professionals to execute steps of analysis as specified by the TURF framework. For the within-model domain function saturation, D2Refine had 96% saturation, which was 4 percentage points better than OntoMaton and 28 percentage points better than RightField. The manual examination and statistical analysis of the data were conducted for task analysis, and the results demonstrated a significant difference for favorability toward D2Refine ( $P<.001$ ) with a 95% CI. Overall, 17 out of 27 (63%) participants indicated that D2Refine was their favorite of the three options.

**Conclusions:** D2Refine is a useful and promising platform that can help address the emerging needs related to clinical research study data dictionary standardization and harmonization.

(*JMIR Hum Factors* 2018;5(3):e10205) doi:[10.2196/10205](https://doi.org/10.2196/10205)

**KEYWORDS**

usability study; data dictionary; interoperability; electronic health records; usability framework; metadata; standardization

## Introduction

The process of creating and managing interoperable metadata is challenging. Additionally, the use of spreadsheets or simple tabular forms to express and organize metadata definitions is widespread in the research community. The environments with spreadsheets and tabular interfaces are common, simple, and flexible, and familiarity with creating content in them reduces the learning curve considerably. The trend, toward the ability to manage metadata using the simple tabular interface of a

spreadsheet, is also evident from the list of metadata tools identified by Stanford University Libraries [1]. These solutions have nontrivial installation, configuration, and workflow steps to create and manage metadata. The translation of metadata is usually proprietary and does not adhere to a standard format, reducing interoperability. A standard representation of metadata would assist metadata developers to identify a minimal core set of information and help create metadata models with enhanced interoperability and shared semantics. In our ongoing studies [2], D2Refine Workbench [3] (D2Refine for short) is being developed to address these issues and to make the process of



creating metadata easier using a simpler interface and disseminated models with enhanced interoperability. This greatly reduces the complexity, learning curve, and additional documentation and transformation steps that would otherwise be needed to make models usable outside their local context, when shared.

D2Refine is built on top of an open-source solution called OpenRefine [4] (formerly known as Google Refine), which offers a simpler, spreadsheet-like interface. D2Refine leverages the extensible OpenRefine framework to augment customizable services to create terminology bindings for standardization efforts. It extends export mechanism to serialize models into standard formats to persist and share with other metadata developers. Our objective was to design and conduct a usability study, following the proven methodology of the TURF (task, user, representation, and function) framework of electronic health record (EHR) usability [5], to assess the usability and usefulness of D2Refine platform. The TURF framework comprises four analysis areas of focus to facilitate designing and conducting an effective usability study. The TURF framework helps us gauge three aspects of a system: useful (ability to support the work domain), usable (easy to learn, use, and error-tolerant), and user satisfaction (likability). The TURF framework guidelines for function and task analyses were employed in this usability study for D2Refine. The novelty of this usability study was that it selectively adopted the TURF framework's systematic and nonconfounding function and task analyses guidelines for identifying comparable features of an environment against one or more competing environments. The applicability of the TURF framework guidelines saved us time and effort related to setting up our own guidelines and processes; we instead focused on evaluating environments. Two comparable open-source solutions, OntoMaton [6], developed by Investigation, Study, and Assay (ISA) Tools using the ISA framework, and RightField [7], developed by University of Manchester, were selected for side-by-side comparison with D2Refine. The choices of comparable environments came from our own knowledge about these systems and from a list of metadata tools created by Stanford University Libraries [1]. OntoMaton is a set of plugins to Google Sheets that allows users to manage and standardize data dictionaries, whereas RightField offers similar capabilities through a Java library programmed to work with Microsoft Excel. RightField uses Apache POI Library [8] and Protégé Web Ontology Language (OWL) API [9] to work with spreadsheets and ontologies, respectively. Since all three environments present very similar interfaces (spreadsheet or spreadsheet-like) to clinical study developers, the user analysis and representation analysis aspects of the TURF framework were deferred and not included in this usability study.

This paper describes the requirements collection, execution, and results of the usability study. It includes the selection and results of metrics for function analysis and quantitative evaluation through task analysis. While the function analysis provides insights into the usefulness of D2Refine, the task analysis throws light on the usability of D2Refine for a selected set of tasks. The task analysis is extended to quantify the

satisfaction level of the participants, who completed the selected set of tasks.

## Methods

### Study Design

D2Refine, OntoMaton, and RightField offer viewing, standardizing, and serializing capabilities for data dictionaries with their simple tabular interfaces. In the following subsections, we briefly introduce them and describe their comparable capabilities that are included in the usability study. The TURF framework is the guiding element in this usability study and as we introduce the relevant aspects of the framework. We also describe the participants, who are the most important part of this study.

### Materials

#### *D2Refine Workbench*

As mentioned above, D2Refine is developed by extending an open-source platform, OpenRefine, to help clean up and organize data in an intuitive manner to anyone who is conversant working with spreadsheets. This greatly reduces the learning curve as D2Refine allows a user to create data dictionaries simply by arranging data dictionary variable definitions as rows. In addition, D2Refine leverages the OpenRefine's capability to directly import and ingest content directly from Web-based data dictionaries such as those from the database of Genotypes and Phenotypes (dbGaP) [10]. Figure 1 shows a data dictionary in D2Refine that is imported directly from dbGaP (by using the web address of the data dictionary). The dbGaP metadata elements are marked to demonstrate how simply D2Refine processes and represents them to the user.

D2Refine further extends the built-in reconciliation service mechanism of OpenRefine to standardize the data dictionary variables. D2Refine can add and utilize any Common Terminology Services 2 (CTS2) [11]-compliant terminology service to search and link terms to the data dictionary variable definitions. The D2Refine workbench comes preconfigured with a default reconciliation service, the National Cancer Institute's (NCI) Lexical Enterprise Vocabulary System (LexEVS) CTS2 Service [12], which provides a quick-start for users to standardize the data dictionary content. D2Refine's export or import extensions provide a way to serialize content to a desirable standard or customized format. Although D2Refine implements an extension to serialize a data dictionary to openEHR's Archetype Definition Language (ADL) [13] format, its evaluation for usability was not included in this study.

#### *OntoMaton*

The OntoMaton Google Widget, developed by ISA-Tools [14], is a plug-in widget that works with Google Spreadsheet [15] documents. Once this widget is augmented to Google Spreadsheet as an add-on, it can be invoked to display a right-hand side panel that shares screen with the spreadsheet. OntoMaton lets users connect and search biomedical terms from the National Center for Biomedical Ontology's Bioportal [16,17], Linked Open Vocabularies [18], and Ontology Lookup Service [19].

OntoMaton allows searching with key-phrases in individual and batch mode and organizes by grouping the result candidates, in an effort to help users select an appropriate match and create a terminology (term) binding (linking a cell value to a reference of a controlled vocabulary term; Figure 2). These term bindings are stored as additional worksheets as part of the user’s original data dictionary definition spreadsheet.

**RightField**

RightField is another open-source dataset annotation tool developed by the University of Manchester [20] and the Hiedelberg Institute for Theoretical Studies [21] (Figure 3). It is designed to work as a standalone Java application, which uses Apache’s POI [8] to manage data dictionary content in and as Microsoft Excel documents. Similar to OntoMaton, RightField

opens up a right-hand side panel with a selected set of ontologies to search and select. A user can create or open an existing Microsoft Excel document to work with RightField and standardize the content. RightField is programmed to load multiple ontologies to work with, although it slows down the lookup and sometimes adversely alters the ontology hierarchy tree rendering.

The RightField Ontology Term Annotator allows user to create term bindings for a cell value with a selected matched term from the ontology (illustrated in Figure 3 with a dashed arrow). RightField lets users store a term binding along with a constraint to validate instance data, by allowing the choice of either a class hierarchy or an instance of the matched ontology term. RightField, similar to OntoMaton, manages the term bindings by augmenting auxiliary worksheets to the user’s spreadsheet.

**Figure 1.** A dbGaP Data Dictionary viewed in D2Refine. The column headers describe the metadata elements (variable id, name, type, etc) of a data dictionary. dbGaP: database of Genotypes and Phenotypes.

ID	value	name	type	description	min	max	unit
1. phv00225190.v1	Male	sex	encoded value	Sex of participant	1	2	
2. phv00225191.v1	Female						
3. phv00225192.v1	Missing	agecat	encoded value	Age in 5 year categories	1	14	Years
	15-29						
	30-34						
	35-39						
	40-44						
	45-49						
	50-54						
	55-59						
	60-64						
	65-69						

**Figure 2.** The OntoMaton Ontology Annotator works with Google Spreadsheets as an add-on to a user’s Google Spreadsheets account.

variable-id	variable-name	variable-type	variable-unit	variable-description	variable-default-v
A110	SUBJID	Integer		De-identified subject	
A111	SEX	String, Encoded		Patient's Gender	
A112	Age	Integer	Years	Age of the subject	

Figure 3. A term binding in the RightField Ontology Term Annotator.

	A	B	C	D	E	F
1	variable-id	variable-name	variable-type	variable-unit	variable-description	variable-default
2	A110	SUBJID	Integer		Deidentified subject	
3	A111	SEX	String, Encoded			
4	A112	Age unit	Integer	Years	Age of the patient	
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						

### The TURF Framework of EHR Usability

The TURF Framework of EHR Usability guided the implementation of a usability study to compare environments and defined the ways of measuring the dimensions of usability for each. We used the TURF framework to assess the usefulness of a system by employing the function saturation metrics to help portray coverage of the useful functional capabilities. The task analysis helped us measure efficiency and robustness in task completion workflows and helped us understand the efforts required to accomplish user goals, error prevention, and recovery and learnability, that is, how usable the system was for doing the work. Additionally, the satisfaction level of users could be captured with surveys, interviews, and questionnaires attached to the various tasks users completed over the course of the study.

### Methods of Usability Study

Following institutional review board approval and participants' provision of informed consent, we enrolled 27 participants. Out of these, 15 participants were clinical study developers and 12 participants were a mix of administration and information technology professionals, who develop and support applications for health sciences research at Mayo Clinic, USA. Most of the study developers who participated already used various applications to create, manage, and disseminate clinical study artifacts. For example, some of them were responsible for creating case report forms [22], which are equivalent to data dictionary definitions. These case report forms are composed together, similar to data dictionaries, to design and conduct

studies for various domains of healthcare research. Many of these workflows included referencing- and linking-controlled terminologies for their list of terms and codes. Enrolled participants were invited and vetted for their knowledge and familiarity with the clinical study data dictionaries. Employing the TURF framework for evaluating the usability of D2Refine hinged on recording and learning from the experiences of the participants as D2Refine was compared with OntoMaton and RightField. This study was a combination of two types of analyses: the function analysis and the task analysis. The data gathered helped us quantitatively identify the usability strengths and weaknesses of one system over another, which made it easier to state our conclusions for each system involved.

### Function Analysis

The idea of the function analysis is based on measuring the usefulness of a system by its implementation of essential functions. Function analysis helps identify the implementation of critical functions, without which a system will fail. One of the initial steps of function analysis is to identify the functions and the work domain under which they fall. The functions fall in at least one of the three categories: (1) Functions that a system implements, (2) Functions that users want, and (3) Functions that actually get used to carry out the tasks to accomplish goals. The TURF framework describes these categories as three models: the Designer Model, User Model, and Activity Model, respectively. The level of usefulness of a system is proportional to the overlapped regions of these three models. The TURF framework recommends organizing the identified functions into

a work domain ontology. The collected functions are further evaluated to filter nonessential functions from critical functions.

**The Questionnaire Design for the Function Analysis**

Awareness about the functions that fell into the three models of the TURF framework yielded creation of an effective and useful work domain ontology. This work domain ontology clearly depicted the functional coverage of participating environments. To catalog the desired functionality (and weigh them against implemented and used functions), we designed a questionnaire for the study participants. This questionnaire queried participant’s existing environments as well as what functionalities they wanted to see implemented in a solution. They were quizzed about ways in which they created, stored, and disseminated dataset definitions and their use of controlled vocabularies. The questionnaire included multiple-choice questions as well as questions for detailed free text responses. A careful capture of information from participants proved useful in listing their problems, expectations, and recommendations for an ideal environment.

**The TURF Metrics**

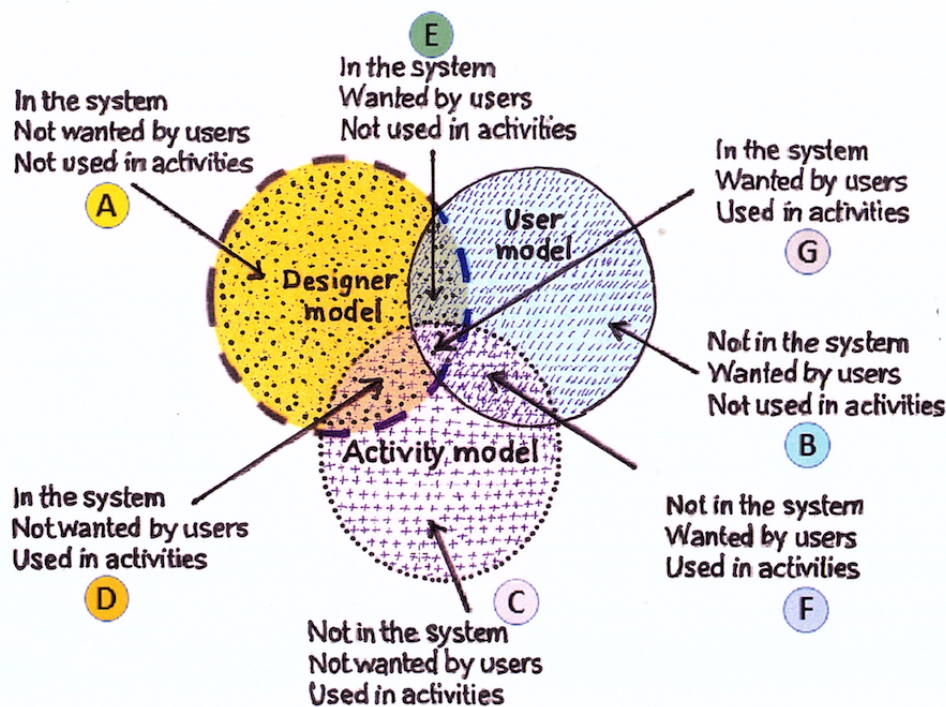
We computed two function saturation metrics of the TURF framework to assess usefulness. The Venn diagrams of TURF framework’s Domain Models [23] in Figure 4 are ways of understanding the coverage (the overlapped areas among the three models) of the functions. The metrics describe the quantitative overlapped portions of implemented functions in an environment (Designer Model) that the user wanted (User Model), and which were eventually used during activities (Activity Model) to accomplish a task.

Within-model domain function saturation: This metric quantifies the functional coverage by the current implementation of functions and is computed as the ratio of domain functions to total functions in the Designer Model. This is calculated as the sum of the number of functions in the regions of D, E, and G divided by the sum of the number of functions in the regions of A, D, E, and G (Figure 4). The number of functions in regions A will be inversely proportional to the domain function saturation or coverage.

Across-model domain function saturation: This is the ratio of domain functions in the Designer Model to the domain functions in all three models (Designer, User, and Activity). It is calculated as the sum of the numerators of the fractions in the regions of A, D, E, and G divided by the count of all domain functions in all three models (Figure 4).

The user questionnaire, discussed in the previous section, was instrumental in capturing the domain functions in these three models. An OWL [24] work domain ontology was created to persist and gain statistical insight into the list of uniquely identified functions. This ontology had a number of top-level OWL classes to partition functions into three models: Design, User, and Activity. Each identified function descended from these top-level classes and each instance of such class related to its implementation in one or more environments. Arranging the classes and instances this way allowed us to quickly and programmatically compute TURF metrics for functional coverage. A set of scripts in Python [25] were developed to query the work domain ontology by using SPARQL Protocol and RDF Query Language [26]. These utility scripts were also used to dynamically create Venn diagrams to illustrate results for better understanding.

**Figure 4.** The Venn diagrams of the TURF (task, user, representation, and function) Framework’s domain models are labeled to help understand the domain function saturation metrics computation.



### Task Analysis

The TURF framework describes task analysis as the process of identifying the steps (mental and physical) and their interdependence to carry out an operation by using a specific representation in an environment. In this usability study, the task analysis steps that the participants followed were relatively simple and straightforward. Since all three environments used very similar spreadsheet-like interfaces, the task analysis focused on fair comparison of performing few selected tasks in each of the environments. The order of selecting the environments to perform tasks was random to avoid bias. Each participant was provided the introduction and appropriate documentation to the three testing environments, D2Refine, OntoMaton, and RightField, in a separate tutorial session. The tutorial sessions were conducted to make participants familiar, comfortable, and conversant in the performance of tasks to accomplish goals of creating, editing, persisting, and standardizing data dictionary elements.

In addition to providing verbal feedback, the participants answered two questions for each task performed in each of the environments. These two questions captured their level of satisfaction using a 5-Point Likert scale [27] (strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree). Although the time taken to perform each task was also recorded, it was not used in weighing one environment's superiority over another to avoid it being a confounding factor. The third element of the task analysis was a survey, one for each environment, which recorded the users' overall experiences.

### The Task Design for the Task Analysis

The participants were given three identical tasks to perform, common to each of the environments: the task of creating and viewing a data dictionary, editing an existing (the newly created dictionary in the first step) data dictionary, and standardizing the variable names of the data dictionary. Aiming for fair comparison, each task had an identical number of steps and method of performing the tasks, and the participants were instructed to follow these steps precisely. At the end of each task, participants were instructed to save the work and check for any loss of work during the save operation.

The first task of creating a data dictionary instructed them to start with an empty data dictionary and then add three variables and their constraint definitions. The second task of editing the data dictionary involved re-opening the data dictionary and adding and editing select variables. The third task (Textbox 1) was to use the variable name as a search key-phrase to search for a matching term from a controlled vocabulary. Each environment had a way of executing the search (block search for all variable names as well as searching for each variable name separately). Once the search result set was presented, each participant chose the best match and created a link between the variable name and a reference to the controlled vocabulary term. These steps made sure an informed valid term binding was created for each variable name.

The participants answered two questions at the end of each task to gauge if the environment allowed them to accomplish the

task well and the satisfaction associated with accomplishing the task, according to their perceived understanding and expectation of the goal. Each question was measured using a 5-Point Likert scale [28] with increasing scores: strongly disagree (1, lowest score), disagree (2), neither agree or disagree (3), agree (4), and strongly agree (5, highest score). Question III was originally designed as part of the representation analysis, which was excluded from the scope of this usability study. It was designed to capture the level of difference between what each user expected from the system and what the user actually did to accomplish the goal. The results were computed with a focus on responses to Question II, rather than Question III.

### Survey Design for the Task Analysis

In an effort to capture additional data that included a participant's overall experience with an environment, we included a set of survey questions (Textbox 2) related to the organization of interface elements, robustness (failure and recovering from a failure, eg, error messages and navigation), auxiliary user interface elements, and easiness with which information about next step could be found.

Survey questions also included overall satisfaction and comfort in using the environments. Each of the 9 survey questions had a binary response: Yes or No. An absence of response was counted as a *No*. Each survey question was also assigned a weight to compute the weighted average score for each environment, in addition to an unweighted average score. At the end, each participant was asked to pick their favorite environment and to rank the environments as first, second, and third in the order of their preference.

The participants were encouraged to provide feedback and convey their experiences, which were recorded as comments. These comments were cataloged and provided useful insights and much sought-after features expected from environments to manage metadata definitions.

### Statistical Methods

#### Function Analysis

The TURF framework metrics (domain function saturation: within-model and across-model) were calculated by using the proportions of functions to demonstrate the functional coverage of each environment. These metrics provide instant understanding of critical and overhead functions each environment implements.

#### Task Analysis

The Kruskal-Wallis test [27], which is a nonparametric method of testing, was chosen to perform the Analysis of Variance [29] for ranking the environments. This method was employed because we did not want to assume normality and our sample size was marginal for parametric testing. A significant Kruskal-Wallis test indicates that an environment is significantly different from others, but it does not indicate how it is different (better or worse). For our purposes, the mean score was determined to be adequate as a mark of overall user experience.

**Textbox 1.** Task details of Study Task 3: to standardize a variable using its name.

### 3. Searching & Binding Controlled Terminology Terms

- a) Open the data dictionary updated in previous task.
- b) Search and Link all the cell values of 'Name' column only. This might involve searching for values that do not find matches.
- c) Verify that all four values in of 'Name' column have been linked.
- d) Save (if needed) and Close the data dictionary.
- e) Please write (or tell the study conductor) answers to the following questions.

**Please record your feedback about this task in this particular environment:**

I. Task completion time: \_\_\_\_\_minutes \_\_\_\_\_seconds

II. The system allows me to accomplish the task well?

1. \_\_\_\_\_Strongly disagree
2. \_\_\_\_\_Disagree
3. \_\_\_\_\_Neither agree or disagree
4. \_\_\_\_\_Agree
5. \_\_\_\_\_Strongly agree

III. The system enables me to accomplish the task well, according to my perceived understanding and expectation of the goal?

1. \_\_\_\_\_Strongly disagree
2. \_\_\_\_\_Disagree
3. \_\_\_\_\_Neither agree or disagree
4. \_\_\_\_\_Agree
5. \_\_\_\_\_Strongly agree

**Textbox 2.** A set of survey questions with their assigned weights were given to the participants to assess their overall experience and level of satisfaction in each of the three environments.

- Q1. It was simple to use this system. (0.75)
- Q2. I can complete my work quickly and efficiently with this system. (1.0)
- Q3. It was easy to learn to use this system. (1.0)
- Q4. The information provided with this system is clear. (1.0)
- Q5. The organization of the information on the system screens is clear. (1.0)
- Q6. Whenever I make mistake using the system, I recover easily and quickly. (1.25)
- Q7. It is easy to find information I needed. (1.0)
- Q8. I feel comfortable using this system. (1.5)
- Q9. Overall, I am satisfied with this system. (1.0)

In the task analysis, there were 26 participants, as one of the participants withdrew before we could conduct the task analysis. As there were more than 2 groups (3 environments), it was useful to see contrasts among environments to precisely understand the performance of one environment compared with another.

To assess these pairwise system differences, we subsequently performed the chi-square test [30]. Pairwise chi-square tests were performed on the dichotomous outcomes in the same manner as we employed for the Likert ranked scores.

A scale transformation was performed prior to the statistical testing, on the Likert scale, from (1, 2, 3, 4, 5) to (-1, -2, 0, 1,

2), to use zero as the center response for better understanding of the results, as a response of three was neutral, below being negative and above being positive. In addition to statistical testing, we also describe our findings in a descriptive side-by-side display of responses for each of the questions.

## Results

### Overview

The results of this usability study not only confirm the usefulness of D2Refine over the other environments but also offer useful insights into potential feature requirements. The participants discussed their experiences with existing systems

and workarounds that they had to take to overcome the lack of functionality and to get jobs done.

**Function Analysis Metrics**

After analyzing each participant’s responses to the user questionnaire and filtering the overhead functions, 98 distinct functions were identified. We used Protégé OWL Editor [31] to create the work domain ontology (Figure 5), where these functions were created as OWL classes and instances. The properties of OWL instances assisted storing membership of a function to its domain model and each participant’s responses.

A Python-based utility queried this work domain ontology to calculate the TURF framework metrics and corresponding Venn diagrams to depict the function coverage. Figure 6 shows the function coverage for three environments. There were 91, 92, and 93 distinct functions for the three environments, respectively. There were 10 common functions that were implemented so far and task analysis employed these common functions like creating and editing data dictionaries. D2Refine had the largest number of functions in the overlapped areas of three models, indicating that testers favored its usefulness more, in comparison to OntoMaton and RightField. D2Refine also showed the least number of unused implemented functions. The results show that D2Refine had the most function saturation by

implementing 25 out of 26 critical Designer Model functions compared with OntoMaton (implemented 16 out of 20 Designer Model functions) and RightField (implemented 11 out of 16 Designer Model functions). Both metrics of the TURF framework for the domain function saturation were calculated (Table 1) among the three platforms, where D2Refine had an edge over others.

The domain function saturation metrics also agreed with the observations of function coverage. It shows that overall, D2Refine implemented 28% of all functions, which was 7 percentage points better than OntoMaton and 11 percentage points than RightField. For the Within-Model domain function saturation, D2Refine had 96% saturation, which was 4 percentage points better than OntoMaton and 28 percentage points better than RightField.

Table 1 shows the TURF framework metrics of domain function saturation calculated according to their definitions in the sections “Function Analysis Metrics” and “The TURF Metrics” and Figure 4. D2Refine implemented 25 out of 26 Domain Model Functions, resulting in 96% coverage for Within-Model Function Saturation. The Across-Model Function Saturation was 28%, wherein D2Refine implemented 26 out of the 91 identified functions. Similar metrics were computed for OntoMaton and RightField.

Figure 5. The Work Domain Ontology helped organize and catalog 98 distinctly identified functions.

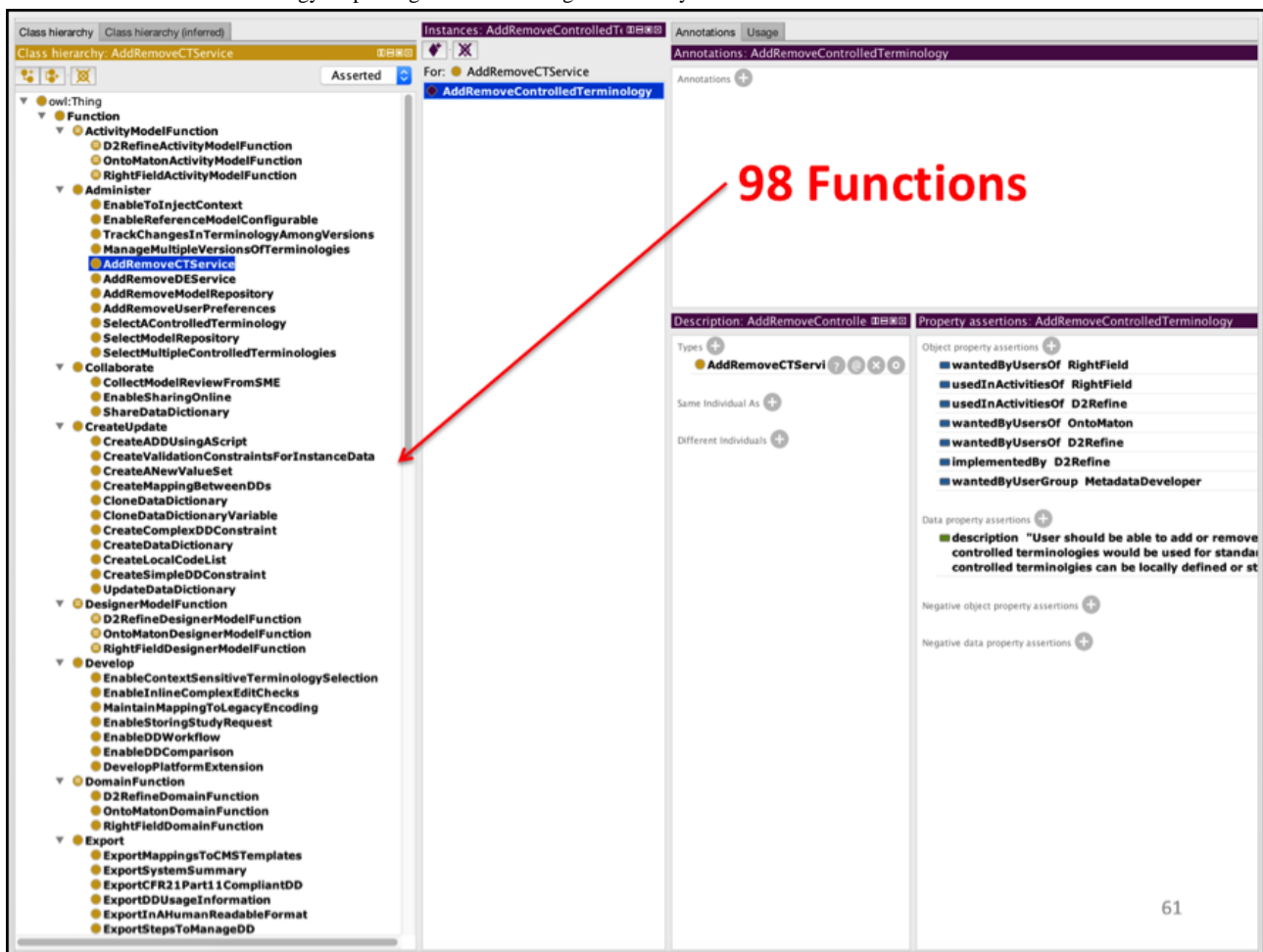


Figure 6. Venn diagrams of function coverage.

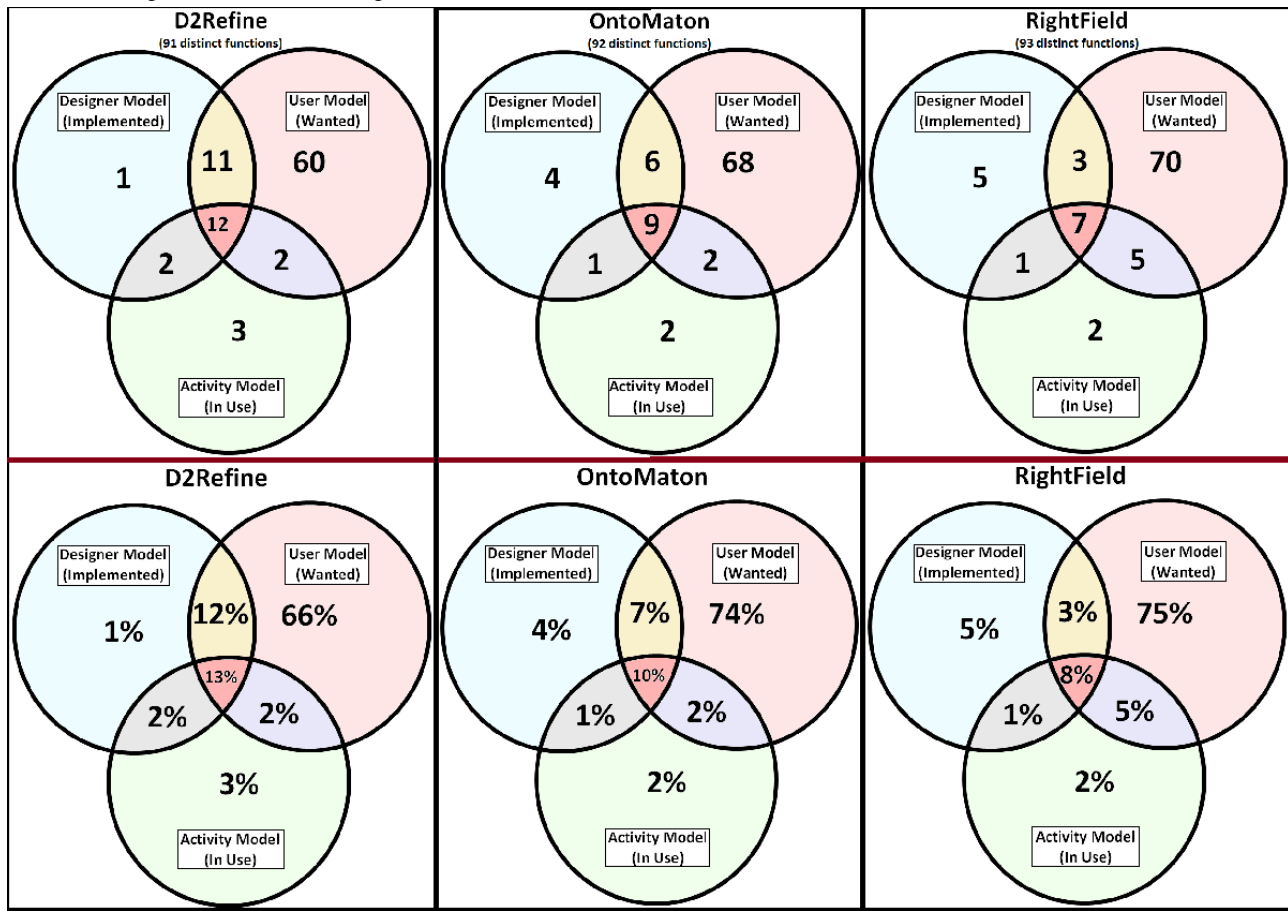


Table 1. The TURF (task, user, representation, and function) framework metrics of domain function saturation calculated according to their definitions.

Domain function saturation	D2Refine		OntoMaton		RightField	
	N	n (%)	N	n (%)	N	n (%)
Within-model	26	25 (96)	20	16 (80)	16	11 (68)
Across-model	91	26 (28)	92	20 (21)	93	16 (17)

**Task Analysis Results**

The results for the task analysis are shown in Tables 2 and 3. Table 2 shows when all three environments were compared for the three identical tasks and mean of their responses for the two questions were tallied. The statistics show that a statistically significant environment score was most helpful in accomplishing the given task well. Please note that while D2Refine was directly comparable with OntoMaton and RightField for Tasks 1 and 2 (creating and updating a data dictionary), it clearly stood out for Task 3 ( $P < .001$ , Kruskal-Wallis test). Figure 7 shows the participant’s responses to the Likert scale choices for the Task 3 questions, which were the most interesting part of the task analysis. Task 3 was nontrivial in all three environments, as it required searching for a matched term and creation of term binding. Both bar charts for Task 3 reflect the favorability of D2Refine over the other two environments.

Here, Table 2 shows the satisfaction level of participants for all three environments, compared during task analysis. The  $P$  values show significant differences and indicate a clear leaning toward D2Refine, especially for Task 3. Table 3 shows the pairwise comparison of D2Refine with OntoMaton and RightField. We observed significant differences for Task 1 and Task 2, indicating that the participants’ experiences were not significantly different for the tasks of creating and editing data dictionaries. However, the statistics showed strong significant differences for Task 3, the task of searching and linking cell values with controlled terminology terms. In other words, the  $P$  values in Table 3 indicate that D2Refine is significantly different from OntoMaton and RightField. Taking into account the statistical leaning toward D2Refine, as exhibited in Table 2, for Task 3, these significant differences indicate the favorability of D2Refine.



**Table 2.** Satisfaction level comparison of the three environments using the average score on Likert scale.

Task and questions	D2Refine	OntoMaton	RightField	<i>P</i> value <sup>a</sup>
<b>Task 1</b>				
Question 1 <sup>b</sup>	1.3	1.5	0.9	<.118
Question 2 <sup>c</sup>	0.8	1.4	0.7	<.014
<b>Task 2</b>				
Question 1	1.5	1.5	0.8	<.036
Question 2	1.2	1.4	0.6	<.017
<b>Task 3</b>				
Question 1	1.2	-0.1	-1.1	<.001
Question 2	1.1	-0.2	-1.3	<.001

<sup>a</sup>*P* values were calculated using Kruskal-Wallis test for a significance level of .05.

<sup>b</sup>Question 1: The system allows me to accomplish the task well?

<sup>c</sup>Question 2: The system enables me to accomplish the task well, according to my perceived understanding and expectation of the goal?

**Table 3.** Pairwise comparison between D2Refine and OntoMaton and RightField<sup>a</sup>.

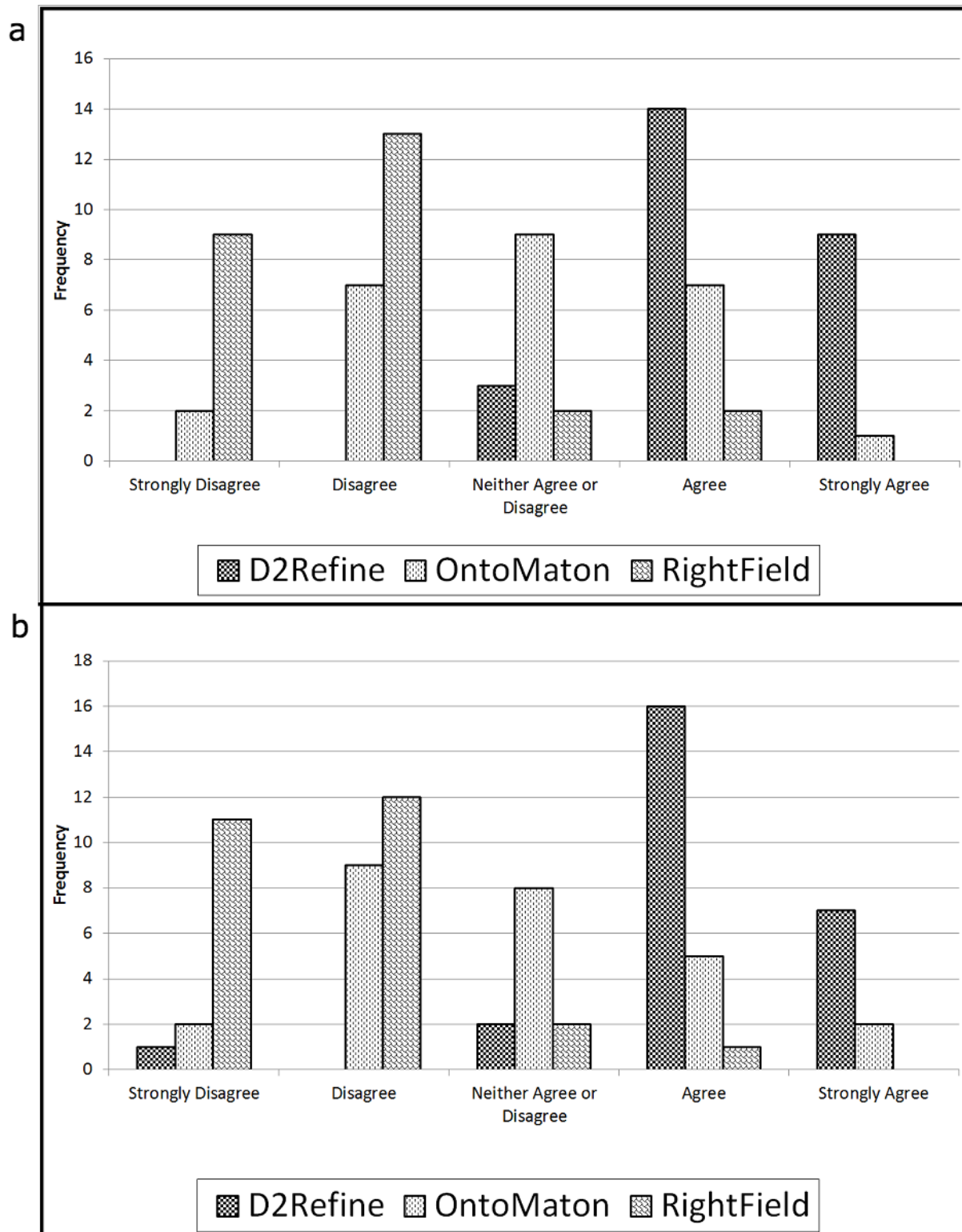
Task and questions	Environment comparison	
	D2Refine vs OntoMaton	D2Refine vs RightField
<b>Task 1</b>		
Question 1 <sup>b</sup>	.05	.19
Question 2 <sup>c</sup>	.023	.56
<b>Task 2</b>		
Question 1	.89	.027
Question 2	.35	.041
<b>Task 3</b>		
Question 1	<.001	<.001
Question 2	<.001	<.001

<sup>a</sup>Statistics show significantly different experience for the tasks (especially Task 3); *P* values were calculated using Kruskal-Wallis test for a significance level of .05.

<sup>b</sup>Question 1: The system allows me to accomplish the task well?

<sup>c</sup>Question 2: The system enables me to accomplish the task well, according to my perceived understanding and expectation of the goal?

**Figure 7.** A side-by-side comparison of the three environments for Task 3 of searching a term and creating terminology binding for a variable name, showing a favorable trend for D2Refine: (a) The system allows me to accomplish the task well? and (b) The system enables me to accomplish the task well, according to my perceived understanding and expectation of the goal?



**Task Analysis Survey Results**

The responses for the 9 survey questions were tallied for their mean scores. We also compared their overall choice for a favorite environment, if it were to be used by participants on regular basis. These results of the survey questions are listed in Table 4. Even though some statistics did not show significant differences (especially for Task 1 and Task 2) among the three

environments, participants showed a strong leaning toward D2Refine for using it to create, update, and standardize the data dictionaries and other metadata creation needs. Here, we demonstrate significant differences with highly significant *P* values ( $P < .001$ ) using the Kruskal-Wallis test for survey sum scores (weighted and nonweighted) and chi-square test for categorical choice for a favorite system.

**Table 4.** Side-by-side comparison of the three environments for the survey results (weighted and nonweighted) and for overall favorite environment.

Measure	D2Refine	OntoMaton	RightField	P value <sup>a</sup>
Nonweighted survey (mean score)	5.4	3.8	1.1	<.001 <sup>b</sup>
Weighted survey (mean score)	5.8	4.1	1.1	<.001 <sup>b</sup>
Participant's favorite environment (participants)	17	9	0	<.001 <sup>c</sup>

<sup>a</sup>P values were calculated using Kruskal-Wallis and chi-square tests for a significance level of .05.

<sup>b</sup>Kruskal-Wallis test.

<sup>c</sup>Chi-square test.

## Discussion

### Principal Findings

In the first prototype of D2Refine, we extended the OpenRefine platform to create and import data dictionaries and standardize them using a CTS2 Reconciliation Service. The aim was to determine D2Refine's usability and effectiveness in managing data dictionaries. Our approach of conducting a moderated usability study in which D2Refine was compared with similar solutions was subsequently clarified. The TURF Framework of EHR Usability was a great tool for designing and planning this usability study. The participants were recruited from a group of interested individuals and included study developers, administrators, information technology professionals, or end users. All had adequate domain knowledge and familiarity with data dictionaries. The tutorials and introduction to these environments were carefully created to avoid favoring a particular environment. This helped us in reducing learning curve greatly for participants and minimizing any confounding factor like lack of domain knowledge.

The interface and workflow steps are almost identical in these three systems for obtaining a data dictionary, updating it, and standardizing its variables. While OntoMaton and RightField leverage the capabilities of Google Spreadsheets and Microsoft Excel, respectively, D2Refine leverages OpenRefine. Participants were given identical or equivalent empty spreadsheets to help carry out the steps. The structured and unstructured responses were gathered and used to calculate TURF metrics and perform statistical side-by-side comparisons.

This usability study provided much needed feedback and insight into the usefulness of D2Refine. The TURF Framework of EHR Usability proved to be a great tool to evaluate the usefulness of each of the participating environments. The function analysis questionnaire helped develop the work domain ontology and also identified 98 distinct functions for possible implementation. Function analysis metrics demonstrated significantly better function coverage (both within and across domain) for D2Refine, as compared to OntoMaton and RightField. Task analysis showed favorable significant differences for accomplishing the identified tasks with D2Refine, especially for term search create terminology binding. Participants' feedback to survey questions and overall experiences favored D2Refine.

### Limitations

While the participants were able to complete their work in all three environments, there were some issues and errors participants faced and some issues translated into recommendations for future improvements. We have highlighted the participants' observations, complaints, and wish-list items that were captured during the task analysis.

The process of typing in the variable definitions was relatively easier in OntoMaton and RightField because participants were directly working with the actual spreadsheets, whereas the D2Refine interface required additional clicks to navigate from one cell to another. Additional steps were needed to add blank rows for new variables in D2Refine. The participants noticed that none of the environments validated the values as they were typed in. The integrated metadata elements (to create and edit data dictionary) of D2Refine platform confused participants with the data dictionary metadata.

For some participants, OntoMaton failed to query and retrieve results, and for some searches, result set categorization was incorrect. Preserving and presenting the term binding details were confusing for OntoMaton and RightField environments. There was no guidance for users to make informed choices when creating term bindings, especially in the case of OntoMaton and RightField. In the case of D2Refine, users could see the term details from reusing the reconciliation service, but this D2Refine functionality could be improved.

The behaviors of interface elements of RightField were disappointing. The column width and font size were very small and cell values were lost due to nonstandard or incorrect interface implementation. There were numerous issues with loading multiple ontologies in RightField and working with them. RightField failed to load moderate to large size ontology, and partial load forced us to reset the working environment and resulted in lost work. RightField always lost the term binding details when data dictionaries reopened and hence heavily discouraged its use.

Although we selected common functionalities for comparison, there are other capabilities that each environment offered, in their own way. We did not include these additional capabilities in this study because they were not common across all three environments. However, two additional features of D2Refine (1) configurable CTS2 Reconciliation Service and (2) serialization of data dictionary into a standard format like openEHR ADL are worth mentioning here.

D2Refine has a built-in reconciliation service, configured to connect to NCI's LexEVS CTS2 Service, which allows users to search for a term in controlled terminologies at NCI. Although the built-in connection service is similar to what OntoMaton and RightField offer, D2Refine lets users add any additional CTS2 compliant service end-point to its list of available reconciliation services. Although the capability of augmenting the reconciliation service for any CTS2-compliant representational state transfer server is not included in this usability study, D2Refine still proves its worth with its built-in reconciliation service, which is at least on-par with OntoMaton and RightField. D2Refine can also persist a data dictionary by serializing it into a standard format such as openEHR ADL [32], which enhances interoperability and makes it shareable and reusable.

Note that installing and configuring OpenRefine for D2Refine, like any other application, requires an additional step that users have to take before D2Refine can be used. This additional step might hinder D2Refine's reach to a wider community, and hence, it forces us to replicate and integrate D2Refine into existing environments. At present, the environments of Microsoft Excel, iMedidata RAVE [33], and SAS Data Management Software [34] are the top choices of participants for starting and working with data dictionaries. These

participants indicated their desire to extend these existing environments to avail the features of D2Refine. As a stand-alone application, D2Refine would still be greatly helpful as a complementary solution to ease the process of study design.

## Conclusions

The benefits of D2Refine's simpler interface and reconciliation feature were validated by this usability study. Even though D2Refine is a prototype for performing data dictionary management, it compares favorably with other existing platforms and environments, which have been evolving over the recent years. The results of this usability study show clear interest and favorability toward the D2Refine platform. Participants not only wanted to see it develop but also to use it as an auxiliary solution that complements their work environment. This usability study provided valuable data as we evaluated our strategy for D2Refine and informed the improvement areas for future development. We believe that the outcome of this work will significantly improve the capabilities of existing informatics tools to manage heterogeneous clinical study data dictionaries and their standardization to improve semantic interoperability of the resulting data models. The artifacts including questionnaires, work domain ontology, and Python utility produced in this study are available online [35].

## Acknowledgments

This paper is extended from a podium abstract presented at the American Medical Informatics Association Annual Symposium 2017. This work was supported in part by funding from an NIH R01 GM105688 and an NCI U01 Project – caCDE-QA (U01 CA180940).

## Conflicts of Interest

None declared.

## References

1. Stanford Libraries. Metadata Tools URL: <http://library.stanford.edu/research/data-management-services/data-best-practices/creating-metadata/metadata-tools> [accessed 2018-04-30] [WebCite Cache ID 6z4YIQQIW]
2. Sharma DK, Solbrig HR, Prud'hommeaux E, Lee K, Pathak J, Jiang G. D2Refine: A Platform for Clinical Research Study Data Element Harmonization and Standardization. AMIA Jt Summits Transl Sci Proc 2017;2017:259-267 [FREE Full text] [Medline: 28815140]
3. GitHub. D2Refine URL: <https://github.com/caCDE-QA/D2Refine> [accessed 2018-04-30] [WebCite Cache ID 6z4YIIQV9]
4. OpenRefine. OpenRefine URL: <http://openrefine.org/index.html> [accessed 2018-07-11] [WebCite Cache ID 6z5n0mnNq]
5. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011 Dec;44(6):1056-1067 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.005] [Medline: 21867774]
6. Maguire E, González-Beltrán A, Whetzel PL, Sansone S, Rocca-Serra P. OntoMaton: a bioportal powered ontology widget for Google Spreadsheets. Bioinformatics 2013 Feb 15;29(4):525-527 [FREE Full text] [doi: 10.1093/bioinformatics/bts718] [Medline: 23267176]
7. Wolstencroft K, Owen S, Horridge M, Krebs O, Mueller W, Snoep JL, et al. RightField: embedding ontology annotation in spreadsheets. Bioinformatics 2011 Jul 15;27(14):2021-2022. [doi: 10.1093/bioinformatics/btr312] [Medline: 21622664]
8. The Apache Software Foundation. Apache POI - the Java API for Microsoft Documents URL: <https://poi.apache.org/> [accessed 2018-05-01] [WebCite Cache ID 6z5pxtj21]
9. Protege. Protege-OWL API Programmer's Guide URL: [https://protegewiki.stanford.edu/wiki/ProtegeOWL\\_API\\_Programmers\\_Guide](https://protegewiki.stanford.edu/wiki/ProtegeOWL_API_Programmers_Guide) [accessed 2018-05-01] [WebCite Cache ID 6z65yICJM]
10. NCBI. The database of Genotypes and Phenotypes URL: <https://www.ncbi.nlm.nih.gov/gap> [accessed 2018-07-11] [WebCite Cache ID 70gI8X0oW]
11. Object Management Group. About The Common Terminology Services 2 Specification Version 1.1 URL: <https://www.omg.org/spec/CTS2/1.1/About-CTS2/> [accessed 2018-04-30] [WebCite Cache ID 6z4cVvL2q]

12. NCI Wiki. 2016. LexEVS URL: <https://wiki.nci.nih.gov/display/LexEVS/LexEVS> [accessed 2018-05-01] [WebCite Cache ID 6z5oVIJAm]
13. OpenEHR. Archetype Model (AM) - latest URL: <http://www.openehr.org/releases/AM/latest/docs/> [WebCite Cache ID 6z5oo3aaV]
14. ISA Tools. URL: <http://isa-tools.org/> [accessed 2018-07-11] [WebCite Cache ID 6z5qyOc4u]
15. Google Spreadsheets. URL: <https://www.google.com/sheets/about/> [accessed 2018-05-01] [WebCite Cache ID 6z5rV1Fe1]
16. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009 Jul;37(Web Server issue):W170-W173 [FREE Full text] [doi: 10.1093/nar/gkp440] [Medline: 19483092]
17. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story M, et al. The National Center for Biomedical Ontology. *J Am Med Inform Assoc* 2012;19(2):190-195 [FREE Full text] [doi: 10.1136/amiainl-2011-000523] [Medline: 22081220]
18. Linked Open Vocabularies. URL: <http://lov.okfn.org/dataset/lov> [accessed 2018-05-01] [WebCite Cache ID 6z63DUQDF]
19. EMBL-EBI. Ontology Lookup Service URL: <https://www.ebi.ac.uk/ols/search?q=diabetes> [accessed 2018-05-01] [WebCite Cache ID 6z63mH911]
20. The University of Manchester. URL: <http://www.manchester.ac.uk/> [accessed 2018-05-02] [WebCite Cache ID 6z6atHgev]
21. The Hiedelberg Institute for Theoretical Studies. URL: <https://www.h-its.org/en/> [accessed 2018-05-02] [WebCite Cache ID 6z6bWGjkO]
22. Wikipedia. Case report form URL: [https://en.wikipedia.org/wiki/Case\\_report\\_form](https://en.wikipedia.org/wiki/Case_report_form) [accessed 2018-05-01] [WebCite Cache ID 6z5v2qSDa]
23. Chen J. Developing A Method For Identifying And Reducing Functional Discrepancies Of Information Systems Doctoral dissertation. The University of Texas School of Health Information Sciences: The University of Texas; 2007.
24. World Wide Web Consortium. Owl Web Ontology Language Current Status URL: <http://www.w3.org/standards/techs/owl> [accessed 2018-05-01] [WebCite Cache ID 6z5ysu2se]
25. Python. About Python URL: <https://www.python.org/about/> [accessed 2018-05-01] [WebCite Cache ID 6z60NWGMx]
26. World Wide Web Consortium. SPARQL Query Language for RDF URL: <http://www.w3.org/TR/rdf-sparql-query/> [accessed 2018-05-01] [WebCite Cache ID 6z64HQyfe]
27. Laerd Statistics. Kruskal-Wallis H Test using SPSS Statistics URL: <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php> [accessed 2018-05-01] [WebCite Cache ID 6z64oUA2o]
28. Wikipedia. Likert scale URL: [https://en.wikipedia.org/wiki/Likert\\_scale](https://en.wikipedia.org/wiki/Likert_scale) [accessed 2018-05-31] [WebCite Cache ID 1527794562578077]
29. Wikipedia. Analysis of variance URL: [https://en.wikipedia.org/wiki/Analysis\\_of\\_variance](https://en.wikipedia.org/wiki/Analysis_of_variance) [accessed 2018-05-01] [WebCite Cache ID 6z655gBqQ]
30. Wikipedia. Chi-squared test URL: [https://en.wikipedia.org/wiki/Chi-squared\\_test](https://en.wikipedia.org/wiki/Chi-squared_test) [accessed 2018-05-01] [WebCite Cache ID 6z65TXINX]
31. Stanford Center for Biomedical Informatics Research. Protege Editor URL: <https://protege.stanford.edu/about.php> [accessed 2018-05-01] [WebCite Cache ID 6z65iOGB1]
32. Sharma DK, Solbrig HR, Prud'hommeaux E, Pathak J, Jiang G. Standardized Representation of Clinical Study Data Dictionaries with CIMI Archetypes. *AMIA Annu Symp Proc* 2016;2016:1119-1128 [FREE Full text] [Medline: 28269909]
33. Medidata. URL: <https://www.mdsoft.com/en/products/rave> [accessed 2018-05-01] [WebCite Cache ID 6z66K7eFA]
34. SAS. SAS Data Management URL: [https://www.sas.com/en\\_us/software/data-management.html](https://www.sas.com/en_us/software/data-management.html) [accessed 2018-05-02] [WebCite Cache ID 6z6bu56VS]
35. Github. D2Refine Usability Study URL: <https://github.com/caCDE-QA/D2RefineUsabilityStudy> [accessed 2018-05-01] [WebCite Cache ID 6z66uGaYA]

## Abbreviations

- ADL:** Archetype Definition Language
- CTS2:** Common Terminology Services
- dbGaP:** database of Genotypes and Phenotypes
- EHR:** electronic health record
- ISA:** Investigation, Study, and Assay
- LexEVS:** Lexical Enterprise Vocabulary System
- NCI:** National Cancer Institute
- OWL:** Web Ontology Language
- TURF:** task, user, representation, and function

*Edited by G Eysenbach; submitted 26.02.18; peer-reviewed by L Cui, Z He; comments to author 05.04.18; revised version received 31.05.18; accepted 16.06.18; published 25.07.18.*

*Please cite as:*

*Sharma DK, Peterson KJ, Hong N, Jiang G*

*The D2Refine Platform for the Standardization of Clinical Research Study Data Dictionaries: Usability Study*

*JMIR Hum Factors 2018;5(3):e10205*

*URL: <http://humanfactors.jmir.org/2018/3/e10205/>*

*doi: [10.2196/10205](https://doi.org/10.2196/10205)*

*PMID: [30045832](https://pubmed.ncbi.nlm.nih.gov/30045832/)*

©Deepak Kumar Sharma, Kevin Jerrold Peterson, Na Hong, Guoqian Jiang. Originally published in JMIR Human Factors (<http://humanfactors.jmir.org>), 25.07.2018. 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 <http://humanfactors.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Software for Administering the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events: Usability Study

Martin W Schoen<sup>1</sup>, MD, MPH; Ethan Basch<sup>2,3,4,5</sup>, MD, MSc; Lori L Hudson<sup>6</sup>, PhD; Arlene E Chung<sup>4,7,8</sup>, MD, MHA, MMCi; Tito R Mendoza<sup>9</sup>, PhD; Sandra A Mitchell<sup>10</sup>, CRNP, PhD; Diane St. Germain<sup>11</sup>, RN, MS; Paul Baumgartner<sup>12</sup>, MS; Laura Sit<sup>3</sup>, MS; Lauren J Rogak<sup>3</sup>, MA; Marwan Shouery<sup>3</sup>, MS; Eve Shalley<sup>13</sup>, BA; Bryce B Reeve<sup>14</sup>, PhD; Maria R Fawzy<sup>15</sup>, MHA; Nrupen A Bhavsar<sup>16</sup>, PhD; Charles Cleeland<sup>9</sup>, PhD; Deborah Schrag<sup>17</sup>, MD, MPH; Amylou C Dueck<sup>18</sup>, PhD; Amy P Abernethy<sup>6,14,19</sup>, MD, PhD

<sup>1</sup>Division of Hematology and Medical Oncology, Department of Internal Medicine, Saint Louis University School of Medicine, Saint Louis, MO, United States

<sup>2</sup>Division of Hematology/Oncology, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, United States

<sup>3</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, United States

<sup>4</sup>Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, United States

<sup>5</sup>Department of Health Policy and Management, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, United States

<sup>6</sup>Duke Clinical Research Institute, Duke University, Durham, NC, United States

<sup>7</sup>Division of General Medicine and Clinical Epidemiology, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, United States

<sup>8</sup>Division of General Pediatrics & Adolescent Medicine, Department of Pediatrics, Program on Health & Clinical Informatics, University of North Carolina School of Medicine, Chapel Hill, NC, United States

<sup>9</sup>Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

<sup>10</sup>Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, United States

<sup>11</sup>Division of Cancer Prevention, National Cancer Institute, Rockville, MD, United States

<sup>12</sup>SemanticBits, Herndon, VA, United States

<sup>13</sup>Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD, United States

<sup>14</sup>Duke Cancer Institute, Durham, NC, United States

<sup>15</sup>FHI 360, Durham, NC, United States

<sup>16</sup>Division of General Internal Medicine, Duke University School of Medicine, Durham, NC, United States

<sup>17</sup>Division of Population Sciences, Dana-Farber Cancer Institute, Boston, MA, United States

<sup>18</sup>Alliance Statistics and Data Center, Mayo Clinic, Scottsdale, AZ, United States

<sup>19</sup>Flatiron Health, New York, NY, United States

**Corresponding Author:**

Ethan Basch, MD, MSc

Division of Hematology/Oncology

Department of Medicine

University of North Carolina School of Medicine

170 Manning Drive

Chapel Hill, NC,

United States

Phone: 1 919 843 3055

Email: [ebasch@med.unc.edu](mailto:ebasch@med.unc.edu)

**Abstract**

**Background:** The US National Cancer Institute (NCI) developed software to gather symptomatic adverse events directly from patients participating in clinical trials. The software administers surveys to patients using items from the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) through Web-based or automated telephone

interfaces and facilitates the management of survey administration and the resultant data by professionals (clinicians and research associates).

**Objective:** The purpose of this study was to iteratively evaluate and improve the usability of the PRO-CTCAE software.

**Methods:** Heuristic evaluation of the software functionality was followed by semistructured, think-aloud protocols in two consecutive rounds of usability testing among patients with cancer, clinicians, and research associates at 3 cancer centers. We conducted testing with patients both in clinics and at home (remotely) for both Web-based and telephone interfaces. Furthermore, we refined the software between rounds and retested.

**Results:** Heuristic evaluation identified deviations from the best practices across 10 standardized categories, which informed initial software improvement. Subsequently, we conducted user-based testing among 169 patients and 47 professionals. Software modifications between rounds addressed identified issues, including difficulty using radio buttons, absence of survey progress indicators, and login problems (for patients) as well as scheduling of patient surveys (for professionals). The initial System Usability Scale (SUS) score for the patient Web-based interface was 86 and 82 ( $P=.22$ ) before and after modifications, respectively, whereas the task completion score was 4.47, which improved to 4.58 ( $P=.39$ ) after modifications. Following modifications for professional users, the SUS scores improved from 71 to 75 ( $P=.47$ ), and the mean task performance improved significantly (4.40 vs 4.02;  $P=.001$ ).

**Conclusions:** Software modifications, informed by rigorous assessment, rendered a usable system, which is currently used in multiple NCI-sponsored multicenter cancer clinical trials.

**Trial Registration:** ClinicalTrials.gov NCT01031641; <https://clinicaltrials.gov/ct2/show/NCT01031641> (Archived by WebCite at <http://www.webcitation.org/708hTjITl>)

(*JMIR Hum Factors* 2018;5(3):e10070) doi:[10.2196/10070](https://doi.org/10.2196/10070)

## KEYWORDS

usability; patient-reported outcomes; symptoms; adverse events; PRO-CTCAE; cancer clinical trials

## Introduction

Symptomatic adverse events (AEs) such as nausea and fatigue are common in cancer clinical trials [1]. Historically, this information has been reported by clinicians using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE), the most commonly used system for AE reporting [2]. To enable patients to directly report this information, the NCI recently developed the Patient-Reported Outcomes version of the CTCAE (PRO-CTCAE) item library as a companion to CTCAE. PRO-CTCAE includes 78 symptomatic AEs; for each symptomatic AE, 1-3 distinct items are used to evaluate the presence, frequency, severity, and associated interference with usual or daily activities for a total of 124 items [3]. PRO-CTCAE is designed to be administered frequently in trials, for example, weekly, and it records the worst magnitude for severity assessment, in accordance with the tenets of AE reporting. These AEs can be individually elicited and are not aggregated into global scores compared with other reporting methods. Development and testing of PRO-CTCAE items, including validity, reliability, responsiveness, mode equivalence, and recall period, have been previously reported [4-6].

As part of the development of the PRO-CTCAE items, prototype software was developed [7]. The key functionalities were derived from an iterative process, including patients, clinical trialists, administrators, NCI, and Food and Drug Administration stakeholders, and included the following:

1. *Professional (clinician and research associate) interface:* This includes a form builder that enables selection of PRO-CTCAE items and a configurable alert system that activates emails if patients miss a scheduled self-report or patients self-report a severe or worsening AE. Additionally,

it includes tools for displaying patient-reported information with various levels of access restriction, given the use of the software by different user types.

2. *Survey scheduling:* A graphical calendar that enables scheduling or timing of patient survey administration, which is configurable by study and has the ability to shift dates in real time at the patient or study level if treatment schedules are modified during a given trial.
3. *Patient survey interface:* Surveys are administered to patients through a Web-based survey that presents questions for each AE together on a page (based on prior research) [8] or an automated telephone interactive voice response (IVR) system. "Conditional branching" is included for AEs with more than one question and a free text box is included at the end for patients to add additional symptoms via dropdown options or to enter unstructured text.

Creating such a system is complex, given the necessity for considerations around security and privacy, diverse computer literacy levels of patients, the need to integrate PRO data into the workflow of professionals, and required compliance with US government Section 508 specifications to ensure that the software was accessible to users with disabilities [9]. Thus, before scaling the system for large-scale implementation in clinical trials, we sought to optimize its usability by testing with end users (patients and clinical trial staff). We have described the usability assessment of the PRO-CTCAE system with a combination of evaluation methods in order to facilitate future adoption of the system into oncology research efforts [10] and improve clinical data collection and patient safety [9].

The aims of this study were (1) to perform a heuristic evaluation of the software to determine functionality problems, deviations from best practice, and compliance with regulations; (2) to



conduct Round 1 of the initial usability testing using both quantitative and qualitative methods with target users, patients with cancer, and professionals that treat cancer; and (3) to refine the PRO-CTCAE system with software development and re-evaluate its usability with Round 2 of testing and include remote testing and IVR system evaluation.

## Methods

### Study Approach

A protocol for usability testing was approved by the institutional review boards at the NCI and 3 participating institutions, Duke University (Durham, NC), MD Anderson Cancer Center (Houston, TX), and Memorial Sloan Kettering Cancer Center (New York, NY). The study approach to test and refine the PRO-CTCAE software consisted of two interrelated components, heuristic evaluation, followed by successive rounds of iterative user-based usability testing (Figure 1), to interrogate the following discrete, well-established domains [10,11]: ease of learning, efficiency of use, memorability, error frequency and severity, and subjective satisfaction (Multimedia Appendix 1) [12].

#### Aim 1: Heuristic Evaluation

Heuristic evaluation is an inspection method that identifies usability problems through examination to evaluate compliance with recognized principles [13]. Usability experts interacted with the system and performed all tasks involved in creating and completing a survey to identify common issues related to collection and communication of PRO-CTCAE data for cancer clinical trials [14]. Usability heuristics were applied to all tasks of both patient and professional users to facilitate patient symptom reporting [15]. Results were organized into heuristic categories and discussed by the research team in order to develop solutions, guide software modifications, and identify potential challenges prior to user-based testing [16].

#### Aims 2 and 3: User-Based Usability Testing

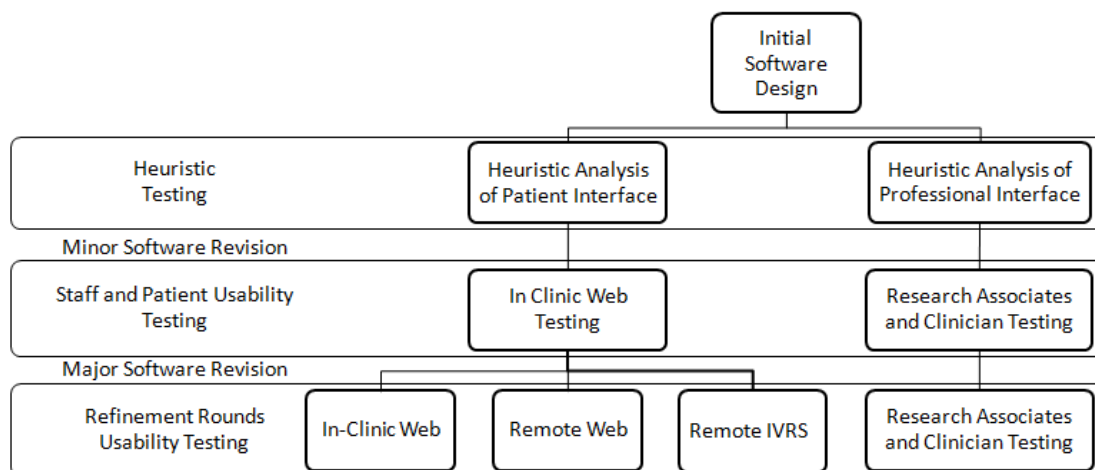
User-based testing involves observation of end users to evaluate the ease of navigation, interaction with application features,

ability to perform essential functions, and satisfaction with task flow [17]. We performed user-based testing of the PRO-CTCAE software with patients receiving systemic cancer treatment and among professional users (physicians, nurses, and research associates). We obtained informed consent from all users for participation in this study.

The usability investigative team (represented by the authors) analyzed the PRO-CTCAE software core functionalities and identified key tasks for testing [18]. The performance of these tasks by end users was directed by experienced evaluators using semiscripted guides that incorporated the “think-aloud” method [19] (see Multimedia Appendix 1 and Patient and Professional Protocols in Multimedia Appendix 2). Evaluators monitored how test subjects interacted with the system, while users were concurrently asked to describe their thoughts and actions during which comments were documented. These comments were categorized into usability problem types and classified as positive, neutral, or negative [20]. We flagged all comments that contained suggestions for improvements for review. Furthermore, a “task completion” scale ranging from 0 to 5 was developed to gauge the difficulty of each usability task (Table 1). After testing, all users completed the System Usability Scale (SUS) that evaluated the usability from 0 to 100, with high scores indicating high usability and scores above 68 indicating better than average usability [21,22].

Consistency among evaluators at each site was emphasized during on-site training conducted by experienced usability evaluators (MS and LH) and was supplemented by subsequent remote booster training. To capture the evaluations of professional staff, evaluators followed a semiscripted guide that was based on prior analysis of key system functions [23]. Accordingly, two rounds of testing were planned, with a targeted sample size of 40-65 professionals (physicians, nurses, and research associates) and 160-195 patients. Based on the conceptual saturation of usability testing issues, the study design included an option to add a third round if usability issues were not resolved through refinement between the first two rounds of testing.

Figure 1. Study description and usability testing plan. IVRS: interactive voice response system (automated telephone).



**Table 1.** Quantitative task completion.

Assistance and score	Score description
<b>Assistance not necessary</b>	
5	Completed task easily
4	Task performed with hesitation or single error
3	Achieved task with confusion or with multiple inappropriate clicks
<b>Assistance provided</b>	
2	Completed with single prompt
1	Task performed after multiple prompts and help
0	Despite prompts, task not completed correctly

The investigative team, in collaboration with human factors consultants, reviewed the results in Round 1 to create solutions and software revision priorities to address the identified limitations in functionality and usability. Subsequently, we tested these modifications in Round 2 and again reviewed to determine if issues were satisfactorily resolved or if further revision or testing was warranted. “Usability” was predefined as the presence of a ceiling effect in the performance measurement and resolution of all identified significant problems amenable to software innovation [19].

### Patient Testing

Patient testing focused on the completion of PRO-CTCAE questions using two different available data entry interfaces, Web-based interface and IVR system. We approached patients receiving outpatient systemic cancer treatment in clinic waiting rooms and invited them to participate in this study if they could speak English and did not have cognitive impairment that would have precluded the understanding of informed consent and meaningful participation in a usability testing. An accrual enrichment strategy was employed to oversample for participants who were ethnically and racially diverse, had high school education or less, were aged >65 years, and had limited baseline computer experience. The accrual of participants with these characteristics was monitored during weekly calls; we discussed strategies for recruiting and enrolling patients with these characteristics.

In Round 1, all participants were asked to perform a series of scripted tasks (eg, log in the system, answer survey questions, and add a symptom), while being observed in private areas of clinic waiting rooms. Evaluators took notes regarding user responses to scripted tasks and questions and audiorecorded the interactions for subsequent transcription and analysis.

In Round 2, patient participants were asked to complete a series of PRO-CTCAE tasks while being monitored in the clinic or remotely without assistance or supervision. For remote testing, patient participants were assigned either to use the Web-based interface or IVR system. Instructions for using these interfaces were provided on an information card with login instructions, and an instructional video was also available. After the remote completion of the PRO-CTCAE tasks, an evaluator contacted each participant and asked semistructured questions about the usability that focused on ease-of-use and difficulties associated with each task. Remote use was emphasized in Round 2 because

it was anticipated that many future trial participants would be accessing the PRO-CTCAE software from home and would not have staff available to assist.

### Physicians, Nurses, and Research Associates Testing

The evaluators observed the users as they completed a scripted series of tasks and audiorecorded encounters for transcription and analysis. In Round 1, the testing was evenly distributed among professional roles, whereas in Round 2, the testing focused predominantly on research associates, as it was anticipated that they would perform a majority of tasks associated with scheduling and processing of PRO-CTCAE data during trials.

### Study Sites

We enrolled all participants from 3 academic cancer hospital outpatient clinics and their affiliated community oncology practices (Duke University, MD Anderson Cancer Center, and Memorial Sloan Kettering Cancer Center). Recruitment was monitored weekly to ensure that the accrual was on schedule and enrichment procedures were being followed and to reinforce consistency of study methods.

### Statistical Analysis

All data were entered into REDCap version 4 (Vanderbilt University, Nashville, TN) SPSS version 21 (IBM, Armonk, NY) was used for analyses. For each usability task, we compared the mean task completion score between each round using independent sample *t* tests and compared them with other tasks in the same round using repeated measures analysis of variance (ANOVA). We performed pairwise comparisons following ANOVA using the Tukey’s Honest Significant Difference test. Furthermore, all statistical tests were two-sided, and we considered  $P < .05$  as statistically significant.

## Results

### Aim 1: Heuristic Evaluation

The system was inspected by 2 usability experts using established heuristics to identify usability issues and propose solutions. Tables 2 and 3 shows the results of this evaluation, including heuristic categories, usability problems, and modifications to the software prior to user-based testing. For example, inspection of a patient Web-based interface revealed that small radio buttons for symptom scoring tended to be

difficult to use by people with poor eyesight and limited dexterity. Thus, the buttons were made larger and the severity of symptom was included in the button (Figure 2). Heuristic and initial patient testing identified difficulty with the use of radio buttons to indicate response choices for symptom collection, lack of apparent progress indicators, and the size,

color, and positioning of navigation buttons (forward and backward) as potential usability issues. Based on these findings, improvements were made to the interface, such as larger buttons, improved indication of button functions, and graphical and numerical progress indicators.

**Table 2.** Results of the heuristic analysis and resulting software solutions (patient).

Heuristic categories	Patient interface issue	Patient interface solution
Visibility of system status	Users cannot tell how many questions remain in a survey	Quantify number of pages or questions remaining and show progress
Match between the system and the real world <sup>a</sup>	Buttons are not representative of their function	Match the shape of buttons to function and add pictures to buttons
User control and freedom	Navigation to move backward and forward not clear	Optimize size, shape, location, and color of forward and backward buttons
Consistency and standards	Inconsistent labeling of PRO-CTCAE <sup>b</sup> symptom terms	Present labeling in consistent format
Error prevention	Small buttons close together, which increases the risk of selecting the wrong button	Increase size, labeling, and spacing of buttons
Aesthetic and minimalist design	Too many radio buttons in variable positions	Enlarge or enhance appropriate buttons; avoid open spaces and scrolling
Help users recognize and recover from errors	No help available	Create help documentation <sup>a</sup>
Disability accommodations <sup>c</sup>	Radio buttons difficult to use; text too small	Create large target area for clicks and touch; make text larger and use easy-to-read font, and appropriate for color-blind individuals

<sup>a</sup>Defined as functionality intuitively matching the intended function.

<sup>b</sup>PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria.

<sup>c</sup>Item is not part of standard traditional heuristics and was added for the specific needs of our patient population.

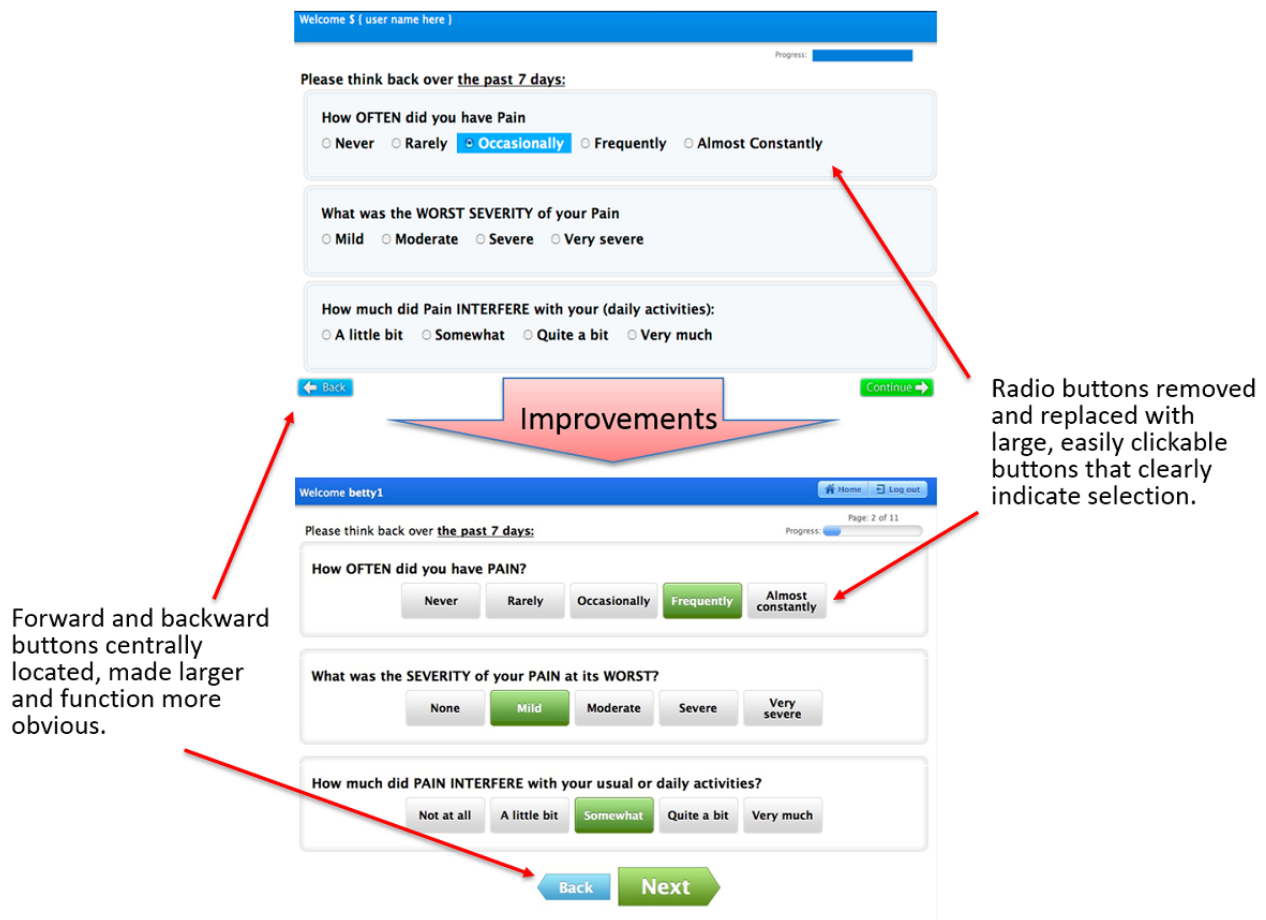
**Table 3.** Results of the heuristic analysis and resulting software solutions (professional users).

Heuristic categories	Professional interface issue	Professional interface solutions
Visibility of system status	Users cannot tell during pauses if the system is processing a task or is frozen	Create a spinning icon to show when the system is processing a task
Match between the system and the real world <sup>a</sup>	Users cannot tell if the survey is ready for patients to complete; survey schedule presented as a list instead of the calendar	Add clear terms for functions (eg, “finalize” to finish a survey); add a graphical calendar to display or alter patient survey schedule
User control and freedom	No ability to customize interface	Provide ability for users to organize interface and modules that they use most often
Consistency and standards	Inconsistent labeling of PRO-CTCAE <sup>b</sup> symptom terms; no ability to download collected data in a standardized format	Present labeling in consistent format; enable data to be downloaded for analysis in common formats
Error prevention	Dates difficult to read, interpret, or change in the survey schedule	Present information in a clear calendar format
Recognition rather than recall	Software does not remember study number or site for a user; the user has to frequently re-enter same data	Software defaults study number and site for users once entered; software auto-populates user preferences or data
Flexibility and efficiency of use	No “dashboard” of essential or time-sensitive data	Create dashboard displaying key information and upcoming surveys
Aesthetic and minimalist design	Menu buttons are difficult to use	Make more functions easily available on the dashboard
Help users recognize and recover from errors	No explanations provided to users to understand causes of errors	Provide popup messages to help correct and prevent future errors

<sup>a</sup>Defined as functionality intuitively matching the intended function.

<sup>b</sup>PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria.

**Figure 2.** Screenshots of the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) before and after usability improvements to benefit end-users: patient.



In the Web-based interface for professional staff, the survey scheduling function was modified from a horizontal list to a more intuitive calendar graphic (Figure 3). Testing with clinicians and research associates identified difficulties in setting and changing schedules for the survey administration to patients. This was improved for Round 2 with the addition of a calendar-type layout with drag-and-drop functionality that enabled survey schedules to be easily configured and modified at the patient level.

## Aim 2: User-Based Usability Assessment

### Participants

Multimedia Appendix 1 shows the characteristics of patients and professionals (physicians, nurses, and research associates) who participated in the usability testing. A total of 169 patients participated; 54.4% (92/169) were females, 18.3% (31/169) were older than 65 years, 31.9% (54/169) were nonwhite, 26.0% (44/169) had high school education or less, and 18.3% (31/169) had limited prior computer experience. Next, 47 professional users participated, including 51% (24/47) clinicians and 49% (23/47) research associates. Clinicians comprised 26% (12/47) physicians and 26% (12/47) nurses.

### Patient Usability Testing: Round 1

Round 1 testing identified favorable initial usability, with a mean SUS score of 86 for the patient Web-based interface (95%

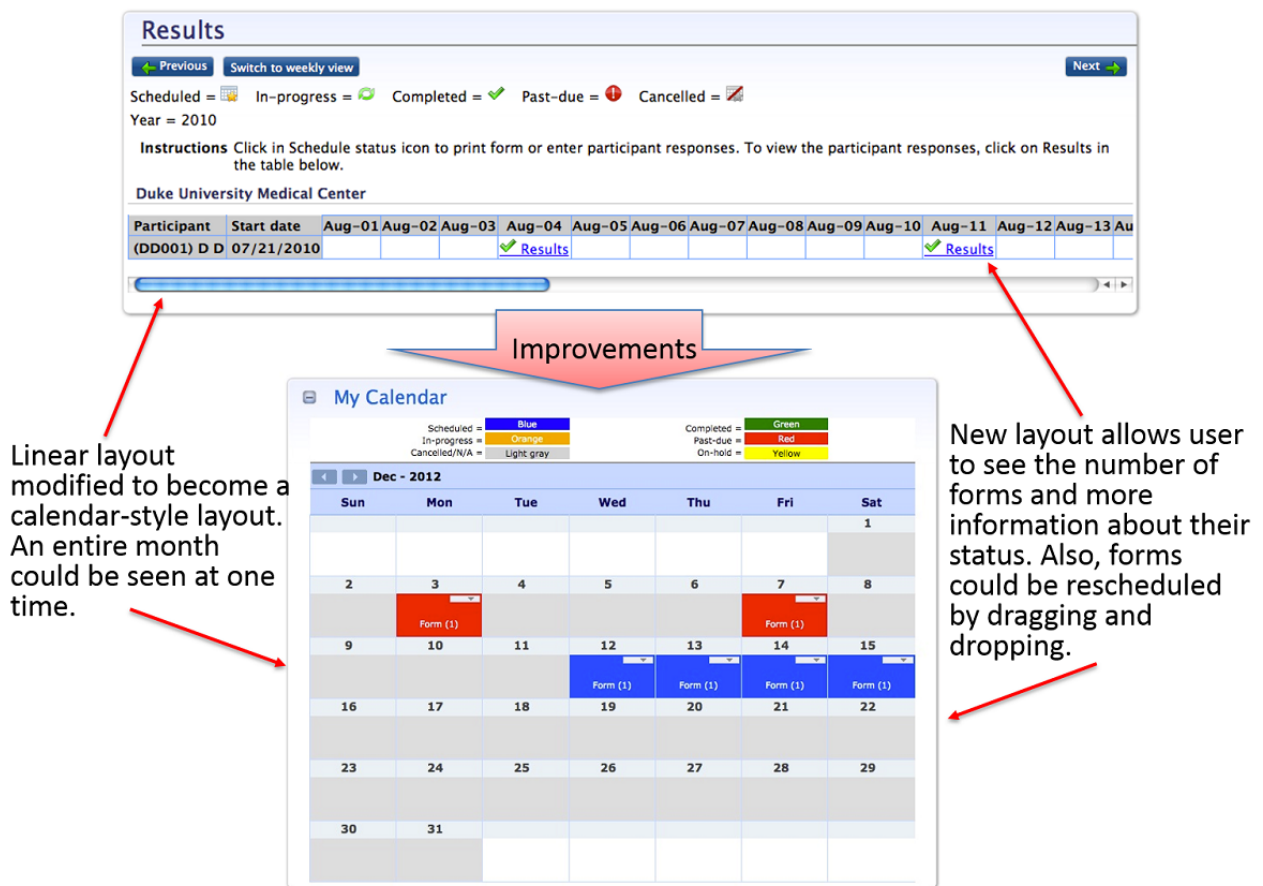
CI 83-90). Figure 4 shows the mean scores for each of the specific tasks using the task completion 0-5 scale. Across all tasks, the mean score was 4.47 (95% CI 4.31-4.62). The only task that was significantly more difficult compared to other tasks was logging into the system (task score 3.67; 95% CI 3.18-4.16;  $P < .001$ ) as shown in Figure 4.

In addition, 51% (90/175) of the comments generated from the think-aloud procedure in Round 1 signified a positive appraisal of the system usability, despite using a protocol designed to find usability problems (Multimedia Appendix 1). Furthermore, 45.1% (79/175) of the patient comments identified areas for improvement, including difficulties with passwords, logging into the system, and problems with the standardized category "match between system and real world" (ie, the task does not intuitively match the intended function).

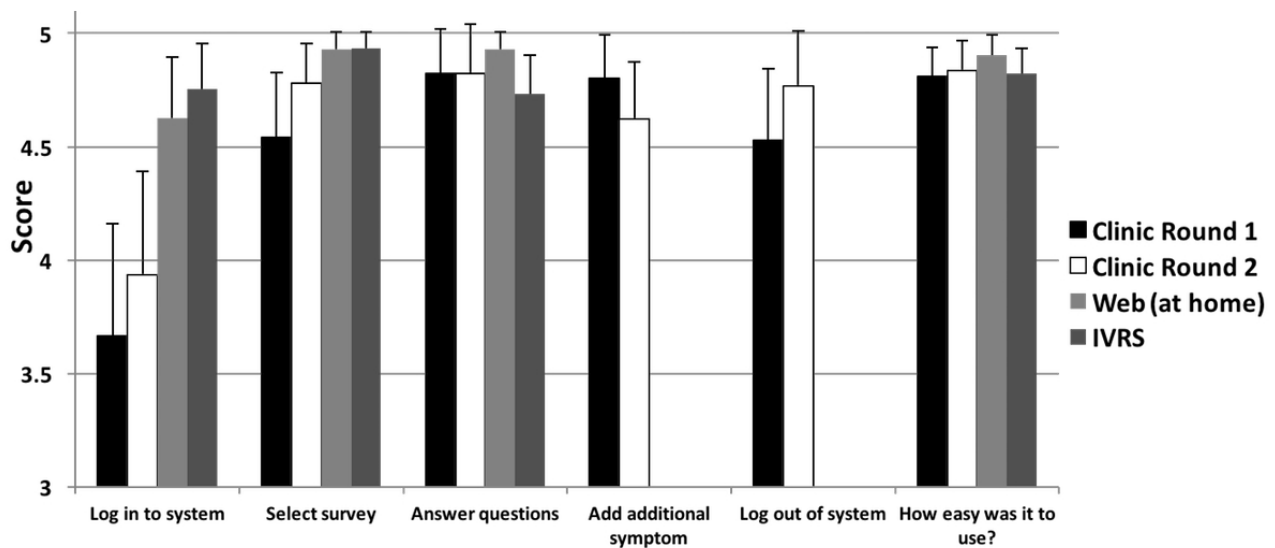
### Professionals (Clinicians and Research Associates) Usability Testing: Round 1

Overall, usability of the system based on the SUS score was 71 (95% CI 60-82). Figure 5 shows the mean task completion score for professional staff users. Moderate to high initial usability was seen across tasks with a mean score of 4.02 (95% CI 3.82-4.21).

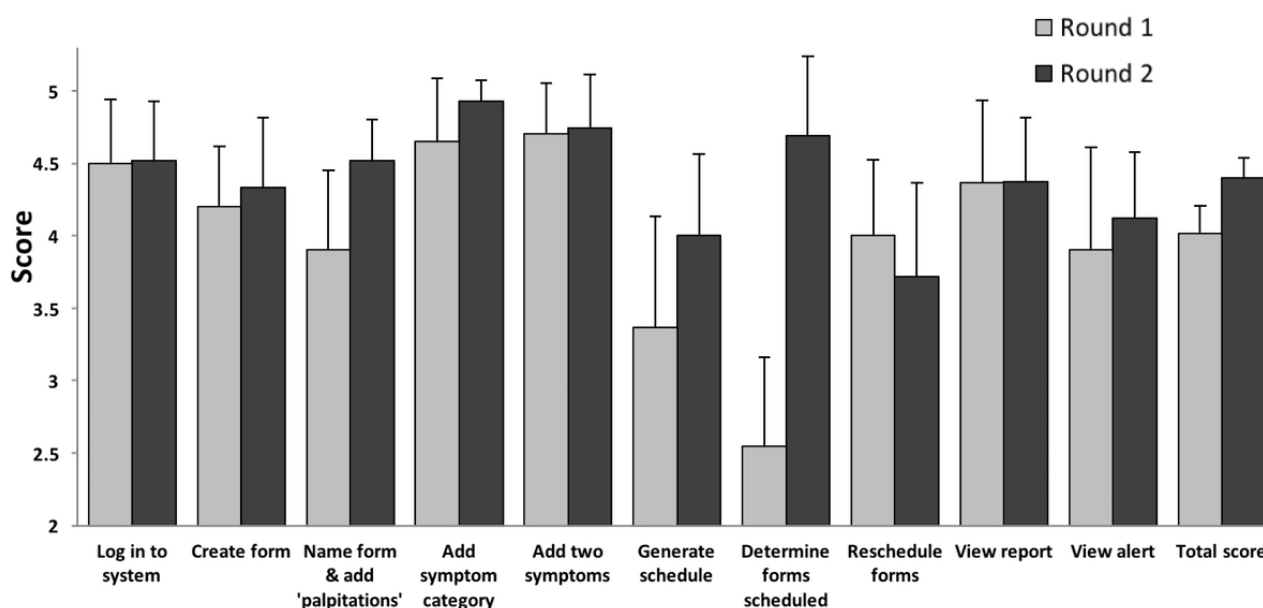
**Figure 3.** Screenshots of the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) before and after usability improvements to benefit end-users: professional staff.



**Figure 4.** Task performance by patients using the 0-5 quantitative task completion scale (Table 1) and tasks from Multimedia Appendix 1. IVRS: interactive voice response system.



**Figure 5.** Task performance by professionals using the 0-5 quantitative task completion scale (Table 1) and tasks from Multimedia Appendix 1.



Several tasks were identified by professional users as difficult or cumbersome, including determining the number of PRO surveys to be administered, monitoring patients' completion of surveys, and creation of a schedule for survey administration. Determining the number of surveys to be scheduled was rated as significantly more difficult than other tasks (task score 2.55; 95% CI 1.90-3.20;  $P=.002$ ), with a trend in task completion score indicating difficulty with scheduling the initial date for survey administration (task score 3.36; 95% CI 2.55-4.19;  $P=.08$ ).

In Round 1 testing, professionals offered 141 comments about system usability and provided recommendations to improve the flexibility and efficiency of use and to provide an aesthetic and minimalist design, recognition rather than recall, a match between the system and the real world (ie, the functionality intuitively matches the intended function), and consistency (Multimedia Appendix 1).

### Aim 3: System Improvements Between Rounds of Testing

Between rounds of user-based testing, software modifications were made based on study results. Specific improvements included functionality for remembering user preferences (eg, defaulting to a user's institution, specific study number or name, and calendar preferences), minimizing the number of required clicks and dialog boxes, and simplifying the design to make the system more intuitive (Figure 3; Multimedia Appendix 1; specific example shown in Figure 2).

A major change to the clinician interface involved the inclusion of a calendar view for PRO-CTCAE survey scheduling. This calendar view could also simultaneously display scheduled surveys for multiple participants on the same day (Figure 3). Other significant changes included the creation of a "dashboard"-type screen upon login, which displayed clinical alerts, upcoming surveys, and the monthly calendar of surveys.

### Patient Usability Testing: Round 2

In Round 2, usability remained high with a mean SUS score of 82 (95% CI 76-88) for the patient Web-based interface as tested in the clinic compared with a mean score of 86 ( $P=.22$  for comparison) in Round 1. Participants who tested the Web-based interface or IVR system remotely and without staff assistance provided mean SUS scores of 92 (95% CI 88-95) for the home Web-based interface and 89 (95% CI 83-95) for the IVR system.

Task completion scores were also high with average score of 4.58 (95% CI 4.45-4.72) for the patient Web-based interface testing in the clinic, 4.85 (95% CI 4.77-4.93) for remote Web testing, and 4.74 (95% CI 4.66-4.82) for remote IVR system testing (Figure 4). Notably, logging into the system continued to be documented as a significant problem when using the patient interface in the clinic where internet connections were inconsistent; the mean score for the task of logging into the system was significantly lower than that for other tasks (3.93; 95% CI 3.46-4.41;  $P=.001$ ). The scores for the remainder of tasks were not found to be markedly different, and the presence of consistently high scores across tasks suggested a ceiling effect.

We analyzed patient user comments separately for clinic-based versus remote use and classified them thematically (Multimedia Appendix 1). The most common theme was "difficulty in logging into the system," which substantially improved between Rounds 1 and 2 (2.3% in Round 2 vs 9.1% of comments in Round 1). The second-most common critique was a lack of "match between the system and real world" (ie, functionality not intuitively matching the intended function), and this mismatch decreased after Round 2 testing (1.7% vs 8.6%). The IVR system component of the PRO-CTCAE system generated negative comments regarding "visibility of system status" (3.7%) and "flexibility and efficiency of use" (2.0%). Based on these results, we concluded that a satisfactory level of patient usability had been attained.

### **Professional Staff Usability Testing: Round 2**

Round 2 testing with professional staff focused on specific usability issues that had been identified in Round 1 and modified through software modifications. In Round 2, the SUS score was 75 (95% CI 69-82), compared with the Round 1 score of 71 ( $P=.47$  for comparison). Across all tasks, the mean task performance score was 4.40 (95% CI 4.26-4.54), which was significantly improved from Round 1 (vs 4.02, 95% CI 3.82-4.21,  $P=.001$ ). Usability scores improved for the 2 tasks with marked difficulty in Round 1, specifically, “determining the number of surveys to be scheduled” (improved from 2.55 in Round 1 to 4.69, 95% CI 4.14-5.24,  $P<.001$ ) and “creating an initial survey administration schedule” (improved from 3.36 in Round 1 to 4.00, 95% CI 3.44-4.56,  $P=.19$ ). Furthermore, the task of “naming a form and adding a symptomatic toxicity” significantly improved from 3.90 in Round 1 to 4.52 (95% CI 4.24-4.80;  $P=.04$ ).

Compared with Round 1, professionals offered fewer negative comments regarding “aesthetic and minimalist design,” as well as “match between the system and the real world” (Multimedia Appendix 1). Negative comments persisted in the heuristic domains of “recognition rather than recall” and “flexibility and efficiency of use.” The investigative team discussed these comments and concluded that they were consistent with the learning curve typically associated with the use of any complex software and that further modifications to address these comments were unlikely to improve usability of the system.

## **Discussion**

A rigorous usability evaluation of a software system for the PRO-CTCAE survey administration, using heuristic and user-based testing with 169 patients and 47 staff members, with iterative modifications between rounds of testing, yielded a highly usable system for electronic capture of PRO-CTCAE responses. As the system for survey scheduling and administration must be integrated into the complex workflow

of cancer clinical trials, comprehensive usability testing by both patients and professional staff was essential. In comparison to many usability assessments, this study included a relatively large and diverse sample that included patients, clinicians, and research associates as users. Moreover, a purposeful enrollment strategy to achieve a patient sample that was diverse with respect to age, ethnicity, educational attainment, and digital literacy strengthens the generalizability of our results.

Based on these favorable usability outcomes, the PRO-CTCAE software system has been implemented in 5 large, multicenter cancer clinical trials in the NCI National Clinical Trials Network and the NCI Community Oncology Research Program (NCORP; NCT01515787, NCT02037529, NCT02414646, NCT01262560, NCT02158637). These findings have also informed the specification of the required functionalities for a downloadable mobile app to collect PRO-CTCAE data within the Medidata Rave clinical trials data management system, thereby supporting the inclusion of PRO-CTCAE in numerous NCI-sponsored cooperative group trials.

This study has several limitations, which should be considered while interpreting the results of this study. First, the system was only assessed in outpatients, and therefore, it is not known whether comparable usability would be seen in hospitalized patients. Second, the sampling did not include participants with visual, auditory, or tactile impairments that might restrict their use of computer hardware or a telephone-based IVR system use of hardware. Finally, we did not enrich our sample for participants with performance status impairment, and approximately 20% of patients enrolled were older than 65 years and had lower digital literacy.

In conclusion, heuristic evaluation followed by iterative rounds of multistakeholder user-based testing and refinement evolved the PRO-CTCAE software into an effective and well-accepted platform for patient-reporting of symptomatic AEs in cancer clinical trials.

## **Acknowledgments**

Funding for this project was provided through contract HHSN261201000043C, awarded by the National Cancer Institute.

## **Authors' Contributions**

The manuscript was authored by MWS and EB with a review from AEC, SAM, ACD, LJR, NAB, and APA. The study was conceived by EMB, BBR, CC, and APA and implemented by MWS, LLH, TRM, SAM, DSG, PB, LS, LJR, MS, MRF, and APA. Statistical analysis was performed by MWS with assistance from TRM and ACD.

## **Conflicts of Interest**

APA is an employee of Flatiron Health, a member of The Roche group.

## **Multimedia Appendix 1**

Information on usability domains used to define the PRO-CTCAE system; tasks included in the evaluation of the patient portal and professional interface; patient user demographics; professional user demographics; patients' qualitative comments about the system; professionals' qualitative comments about the system.

[PDF File (Adobe PDF File), 118KB - [humanfactors\\_v5i3e10070\\_app1.pdf](#) ]

## Multimedia Appendix 2

Professional and patient think-aloud protocols.

[[PDF File \(Adobe PDF File\), 98KB](#) - [humanfactors\\_v5i3e10070\\_app2.pdf](#) ]

### References

1. Reilly CM, Bruner DW, Mitchell SA, Minasian LM, Basch E, Dueck AC, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. *Support Care Cancer* 2013 Jun;21(6):1525-1550 [FREE Full text] [doi: [10.1007/s00520-012-1688-0](#)] [Medline: [23314601](#)]
2. National Cancer Institute, National Institutes of Health. National Cancer Institute Enterprise Vocabulary Services. 2017. Common Terminology Criteria for Adverse Events (CTCAE) URL: [https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE\\_4.03/CTCAE\\_4.03\\_2010-06-14\\_QuickReference\\_5x7.pdf](https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf) [accessed 2018-06-14] [WebCite Cache ID [70AQn01H0](#)]
3. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014 Sep;106(9) [FREE Full text] [doi: [10.1093/jnci/dju244](#)] [Medline: [25265940](#)]
4. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, National Cancer Institute PRO-CTCAE Study Group. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *JAMA Oncol* 2015 Nov;1(8):1051-1059 [FREE Full text] [doi: [10.1001/jamaoncol.2015.2639](#)] [Medline: [26270597](#)]
5. Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, National National Cancer Institute PRO-CTCAE Study Group. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Health Qual Life Outcomes* 2016 Feb 19;14:24 [FREE Full text] [doi: [10.1186/s12955-016-0426-6](#)] [Medline: [26892667](#)]
6. Mendoza TR, Dueck AC, Bennett AV, Mitchell SA, Reeve BB, Atkinson TM, et al. Evaluation of different recall periods for the US National Cancer Institute's PRO-CTCAE. *Clin Trials* 2017 Jun;14(3):255-263. [doi: [10.1177/1740774517698645](#)] [Medline: [28545337](#)]
7. Agarwal HG, Baumgartner P, Coffey D, Kumar V, Chilukuri R, Shouery M, et al. Web interface for the patient-reported version of the CTCAE. In: *Web interface for the patient-reported version of the CTCAE*. 2009 Presented at: NCI caBIG Annual Meeting; 2009; Washington, DC.
8. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *J Am Med Inform Assoc* 2007;14(3):264-268 [FREE Full text] [doi: [10.1197/jamia.M2177](#)] [Medline: [17329732](#)]
9. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *J Am Med Inform Assoc* 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: [10.1136/amiajnl-2012-001458](#)] [Medline: [23355463](#)]
10. Horsky J, McColgan K, Pang JE, Melnikas AJ, Linder JA, Schnipper JL, et al. Complementary methods of system usability evaluation: surveys and observations during software design and development cycles. *J Biomed Inform* 2010 Oct;43(5):782-790 [FREE Full text] [doi: [10.1016/j.jbi.2010.05.010](#)] [Medline: [20546936](#)]
11. Yen P, Bakken S. A comparison of usability evaluation methods: heuristic evaluation versus end-user think-aloud protocol - an example from a web-based communication tool for nurse scheduling. *AMIA Annu Symp Proc* 2009 Nov 14;2009:714-718 [FREE Full text] [Medline: [20351946](#)]
12. US Department of Health & Human Services. 2014. Usability Evaluation Basics URL: <https://www.usability.gov/what-and-why/usability-evaluation.html> [accessed 2018-04-19] [WebCite Cache ID [6ynmDJjQ](#)]
13. Nielsen J, Mack R. Usability inspection methods. In: *Usability Inspection Methods*. New York, NY: John Wiley & Sons; 1994.
14. Abernethy AP, Herndon JE, Wheeler JL, Day JM, Hood L, Patwardhan M, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. *J Pain Symptom Manage* 2009 Jun;37(6):1027-1038. [doi: [10.1016/j.jpainsymman.2008.07.011](#)] [Medline: [19394793](#)]
15. Choi B, Drozdetski S, Hackett M, Lu C, Rottenberg C, Yu L, et al. Usability comparison of three clinical trial management systems. *AMIA Annu Symp Proc* 2005:921 [FREE Full text] [Medline: [16779208](#)]
16. Hartson HR, Andre TS, Williges RC. Criteria For Evaluating Usability Evaluation Methods. *International Journal of Human-Computer Interaction* 2003 Feb;15(1):145-181. [doi: [10.1207/s15327590ijhc1501\\_13](#)]
17. Bastien J. Usability testing: a review of some methodological and technical aspects of the method. *Int J Med Inform* 2010 Apr;79(4):e18-e23. [doi: [10.1016/j.ijmedinf.2008.12.004](#)] [Medline: [19345139](#)]
18. Yen P, Bakken S. Review of health information technology usability study methodologies. *J Am Med Inform Assoc* 2012 Jun;19(3):413-422 [FREE Full text] [doi: [10.1136/amiajnl-2010-000020](#)] [Medline: [21828224](#)]



19. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform* 2004 Feb;37(1):56-76. [doi: [10.1016/j.jbi.2004.01.003](https://doi.org/10.1016/j.jbi.2004.01.003)] [Medline: [15016386](https://pubmed.ncbi.nlm.nih.gov/15016386/)]
20. Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. *J Biomed Inform* 2003;36(1-2):23-30 [[FREE Full text](#)] [Medline: [14552844](https://pubmed.ncbi.nlm.nih.gov/14552844/)]
21. Brooke J. SUS - a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdneester B, editors. *Usability Evaluation in Industry*. Bristol, PA, London: Taylor & Francis, Inc; 1996:189-194.
22. Sauro J. SUSified? Little-Known System Usability Scale Facts. *User Experience Magazine* 2011;10(3) [[FREE Full text](#)]
23. Fromme EK, Kenworthy-Heinige T, Hribar M. Developing an easy-to-use tablet computer application for assessing patient-reported outcomes in patients with cancer. *Support Care Cancer* 2011 Jun;19(6):815-822 [[FREE Full text](#)] [doi: [10.1007/s00520-010-0905-y](https://doi.org/10.1007/s00520-010-0905-y)] [Medline: [20512360](https://pubmed.ncbi.nlm.nih.gov/20512360/)]

## Abbreviations

**AE:** adverse event

**ANOVA:** analysis of variance

**CTCAE:** Common Terminology Criteria for Adverse Events

**IVR:** interactive voice response

**IVRS:** interactive voice response system

**NCI:** National Cancer Institute

**PRO-CTCAE:** Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events

**SUS:** System Usability Scale

*Edited by G Eysenbach; submitted 12.02.18; peer-reviewed by B Hirsch, RL Mercieca-Bebber; comments to author 29.03.18; revised version received 26.04.18; accepted 08.05.18; published 16.07.18.*

### *Please cite as:*

*Schoen MW, Basch E, Hudson LL, Chung AE, Mendoza TR, Mitchell SA, St. Germain D, Baumgartner P, Sit L, Rogak LJ, Shouery M, Shalley E, Reeve BB, Fawzy MR, Bhavsar NA, Cleeland C, Schrag D, Dueck AC, Abernethy AP*

*Software for Administering the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events: Usability Study*

*JMIR Hum Factors* 2018;5(3):e10070

URL: <http://humanfactors.jmir.org/2018/3/e10070/>

doi: [10.2196/10070](https://doi.org/10.2196/10070)

PMID: [30012546](https://pubmed.ncbi.nlm.nih.gov/30012546/)

©Martin W Schoen, Ethan Basch, Lori L Hudson, Arlene E Chung, Tito R Mendoza, Sandra A Mitchell, Diane St. Germain, Paul Baumgartner, Laura Sit, Lauren J Rogak, Marwan Shouery, Eve Shalley, Bryce B Reeve, Maria R Fawzy, Nrupen A Bhavsar, Charles Cleeland, Deborah Schrag, Amylou C Dueck, Amy P Abernethy. Originally published in *JMIR Human Factors* (<http://humanfactors.jmir.org>), 16.07.2018. 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 <http://humanfactors.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Designing Emails Aimed at Increasing Family Physicians' Use of a Web-Based Audit and Feedback Tool to Improve Cancer Screening Rates: Cocreation Process

Caroline A Bravo<sup>1</sup>, MSc; Diego Llovet<sup>1,2</sup>, PhD; Holly O Witteman<sup>3,4,5</sup>, PhD; Laura Desveaux<sup>2,6,7</sup>, PhD; Justin Presseau<sup>8,9,10</sup>, PhD; Marianne Saragosa<sup>7,11</sup>, RN, MN; Gratianna Vaisson<sup>4,5</sup>, MSc; Shama Umar<sup>1</sup>, MSc; Jill Tinmouth<sup>1,12,13,14</sup>, MD, PhD; Noah M Ivers<sup>6,7,11,12,15</sup>, MD, PhD

<sup>1</sup>Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada

<sup>2</sup>Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

<sup>3</sup>Department of Family and Emergency Medicine, Faculty of Medicine, Laval University, Quebec City, QC, Canada

<sup>4</sup>Office of Education and Professional Development, Faculty of Medicine, Laval University, Quebec City, QC, Canada

<sup>5</sup>Research Centre of the CHU de Québec, Quebec City, QC, Canada

<sup>6</sup>Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada

<sup>7</sup>Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

<sup>8</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

<sup>9</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

<sup>10</sup>School of Psychology, Faculty of Social Sciences, University of Ottawa, Ottawa, ON, Canada

<sup>11</sup>Family Practice Health Centre, Women's College Hospital, Toronto, ON, Canada

<sup>12</sup>Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

<sup>13</sup>Department of Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

<sup>14</sup>Sunnybrook Research Institute, Toronto, ON, Canada

<sup>15</sup>Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

**Corresponding Author:**

Noah M Ivers, MD, PhD

Institute for Health System Solutions and Virtual Care

Women's College Hospital

76 Grenville Street

Toronto, ON, M5S 1B2

Canada

Phone: 1 416 323 6060

Fax: 1 416 323 6255

Email: [noah.ivers@utoronto.ca](mailto:noah.ivers@utoronto.ca)

## Abstract

**Background:** Providing clinical performance data to health professionals, a process known as audit and feedback, can play an important role in health system improvement. However, audit and feedback tools can only be effective if the targeted health professionals access and actively review their data. Email is used by Cancer Care Ontario, a provincial cancer agency, to promote access to a Web-based audit and feedback tool called the Screening Activity Report (SAR); however, current emails that lack behavior change content have been ineffective at encouraging log-in to the SAR.

**Objective:** The objective of our study was to describe the process and experience of developing email content that incorporates user input and behavior change techniques (BCTs) to promote the use of the SAR among Ontario primary care providers.

**Methods:** Our interdisciplinary research team first identified BCTs shown to be effective in other settings that could be adapted to promote use of the SAR. We then developed draft BCT-informed email content. Next, we conducted cocreation workshops with physicians who had logged in to the SAR more than once over the past year. Participants provided reactions to researcher-developed BCT-informed content and helped to develop an email that they believed would prompt their colleagues to use the SAR. Content from cocreation workshops was brought to focus groups with physicians who had not used the SAR in

the past year. We analyzed notes from the cocreation workshops and focus groups to inform decisions about content. Finally, 8 emails were created to test BCT-informed content in a 2×2×2 factorial randomized experiment.

**Results:** We identified 3 key tensions during the development of the email that required us to balance user input with scientific evidence, organizational policies, and our scientific objectives, which are as follows: conflict between user preference and scientific evidence, privacy constraints around personalizing unencrypted emails with performance data, and using cocreation methods in a study with the objective of developing an email that featured BCT-informed content.

**Conclusions:** Teams tasked with developing content to promote health professional engagement with audit and feedback or other quality improvement tools might consider cocreation processes for developing communications that are informed by both users and BCTs. Teams should be cautious about making decisions solely based on user reactions because what users seem to prefer is not always the same as what works. Furthermore, implementing user recommendations may not always be feasible. Teams may face challenges when using cocreation methods to develop a product with the simultaneous goal of having clearly defined variables to test in later studies. The expected role of users, evidence, and the implementation context all warrant consideration to determine whether and how cocreation methods could help to achieve design and scientific objectives.

(*JMIR Hum Factors* 2018;5(3):e25) doi:[10.2196/humanfactors.9875](https://doi.org/10.2196/humanfactors.9875)

## KEYWORDS

audit and feedback; clinical audit; co-creation; user design; email; behavior change; physician engagement

## Introduction

Audit and feedback is a widely adopted strategy to improve clinical practice guideline adherence [1]. However, audit and feedback tools can only be effective if physicians access them to review their clinical performance data. Physicians who do not see their performance feedback data are not likely to use it for quality improvement [2]. Email is commonly used to communicate with physicians and can be an effective channel for encouraging guideline-recommended care [3] and access to Web-based educational opportunities [4]. Cancer Care Ontario, a provincial government cancer agency, uses email communications to encourage primary care physicians to log in to a Web-based cancer screening tool called the Screening Activity Report (SAR).

The SAR is a Web-based audit and feedback tool aimed at improving cancer screening-guideline adherence among primary care physicians in Ontario, Canada. The SAR contains patient-level information about rostered patients (ie, patients that are enrolled with a physician) who require action regarding cancer screening (eg, due for screening, overdue for screening, and require follow-up to an abnormal result). To promote regular SAR use, Cancer Care Ontario sends monthly emails to family physicians to inform them that the SAR data has been updated. However, the content of these monthly emails appears to be ineffective at promoting access to the SAR; only 2.37% (129/5445), 3.76% (207/5512), and 4.09% (227/5552) of the SAR-registered physicians logged in to the SAR in February, March, and April 2017, respectively [5]. Additionally, in the same period, less than half of recipients opened the email and only about 7% clicked on the embedded link to access the SAR [6]. Thus, there is a need to understand how to develop emails that are more effective at encouraging family physicians to log in and use the audit and feedback tool.

There may be a number of reasons why physicians do not open the monthly SAR email, including feeling overwhelmed by the number of emails received and having limited time to open emails [7]. Another potential reason may be that physicians dismiss these emails owing to a lack of compelling content (eg, benefits to the physician and patient of using the SAR). It may be possible to improve SAR access with content that employs a behavioral science approach to address barriers to SAR use.

In this paper, we describe our process and experience of developing email content involving cocreation workshops and focus group discussions with physician SAR users. The goal of this process was to develop user- and behavior change technique (BCT)-informed email content for a study testing variants of the email [8]. Using BCT classification systems helps to build evidence about which behavior change intervention components work and under what conditions [9].

## Methods

### Design and Setting

We conducted cocreation workshops and focus groups with family physicians in Toronto and Kingston, Ontario, Canada between January and April 2017. Final email products were tested in a 2×2×2 factorial experiment with Ontario physicians registered for the SAR [8].

### Email Development Process

We used an iterative process to develop user-informed email content that operationalizes 3 BCTs. [Textbox 1](#) illustrates the steps involved in our email content development process. Throughout the process, we engaged relevant decision makers at Cancer Care Ontario to review content and provide feedback to ensure that the final products would be implementable.

**Textbox 1.** Email content development process.

Step 1. Interdisciplinary research team selects behavior change techniques (BCTs) and develops draft content.

- Reviewed existing literature on behavior change among physicians [10].
- Identified potential BCTs from Michie et al's BCT taxonomy (v1) [11].
- Held a team creative writing session to develop sample messages inspired by the examples on Michie et al's BCT taxonomy [11].

Outcome: Sample messages that operationalized BCTs to bring to cocreation workshops for critique, refinement, or replacement by participants.

Step 2. Recruit users.

- Used a combination of purposive [12] and convenience sampling [13] to recruit adopter and nonadopters for cocreation workshops and focus groups.
  - Adopter = physician user who had logged in to the SAR more than once within the 12 months prior to our recruitment period.
  - Nonadopter = physician users who had not logged in to the SAR within those 12 months.
  - Definitions were developed in collaboration with Cancer Care Ontario. Cancer screening for any given patient does not need to happen more than once per year; therefore, it is reasonable to check who is overdue annually. However, it is ideal to do so more often to minimize how far overdue a patient would be and to identify patients that require follow-up to an abnormal result. Additionally, many physicians access the SAR once a year to calculate their cancer screening rates for their application to earn an annual Preventive Care Bonus.
- Participants were offered a Can \$200 gift card in appreciation of their time and effort.

Outcome: Nine adopter participants for cocreation workshops. Eleven participants (adopter and nonadopter) for focus groups. No participant took part in more than one cocreation workshop or focus group.

Step 3. Conduct cocreation workshops with adopters.

- Adapted interview guides for workshops from a planning and design tool for codesign workshops [14].
- Held one 2-hour workshop with adopter participants in Toronto, Ontario. Participants provided feedback on BCT-informed messages developed by the research team in step 1 and produced emails with content that they thought would convince their nonuser counterparts to log in to the SAR.
- Audio-recorded workshops and created detailed notes of reactions to content. Research team met to review adopter-generated content and considered the valence (positive or negative) of reactions to each BCT-informed message developed during step 1, the evidence of effectiveness of each BCT, and the potential risks of using messages given any negative reactions.
- Research team developed 2 emails that combined user-generated content and researcher-developed BCT-informed messages to bring to the second workshop.
- Held one 2-hour workshop with adopter participants in Kingston, Ontario. Participants provided feedback on both emails and refined as needed.
- Research team met to review detailed notes from audio recordings, analyze content for underlying BCTs, and review the valence of reactions.

Outcome: Adopter-generated content and feedback to sample messages.

Step 4. Pretest content with adopters and nonadopters.

- Held three 2-hour focus groups in Toronto and Kingston, Ontario with different adopter and nonadopter participants in each.
- Sent emails that were developed and refined by adopters and the research team in step 3 to participants. Participants viewed emails on their personal cell phone during the focus group. Participants provided reactions and refined content as needed.
- Following each session, the research team met to review detailed notes from audio recordings and discuss user-generated content and participant feedback including verbal reactions to the materials and recommendations for content changes. Considerations for content changes included scientific evidence of effectiveness for each BCT, strength and frequency of the reaction (eg, severity of sentiment such as strong distaste or strong liking, number of participants that shared similar reactions), and feasibility of implementing the recommendation.

Outcome: Adopter-generated content and feedback to emails produced in step 3.

Step 5. Finalize products (emails) for testing.

- Developed 8 different email versions using the user-generated and BCT-informed content that was refined and finalized by the end of step 4.
- Variants of email differed by the number of BCTs operationalized, resulting in different word counts and length.

Outcome: Eight email variants for testing 3 BCTs in a 2×2×2 factorial randomized experiment [8].

## Results

### Participants

Although we made an effort to recruit adopters for cocreation workshops and nonadopter participants for focus groups, challenges with recruitment required us to be open to the inclusion of adopter participants in the focus groups. [Table 1](#) and [Table 2](#) show the participant characteristics for cocreation workshops and focus groups.

### Behavior Change Techniques

The full list of BCTs that were initially selected by the research team and considered throughout the project can be found in [Multimedia Appendix 1](#).

The 3 BCTs to be tested as part of the 2×2×2 factorial randomized experiment include the following: anticipated regret, material incentive (behavior), and problem solving. The final operationalizations of each BCT can be found in [Multimedia Appendix 2](#). The final email with all 3 BCTs can be found in the [Multimedia Appendix 3](#).

### Key Tensions and Resulting Changes to Email Content

Over the course of the email development process, we used 3 main sources of input to inform decisions about content,

including scientific literature, SAR users, and Cancer Care Ontario communication policies. Tensions arose when these sources of input conflicted with each other.

### *Conflict Between User Preferences and Scientific Evidence*

Given the evidence that physicians' decisions are influenced by the desire to minimize regret associated with a potentially wrong decision [15-17], we developed anticipated regret content and rendered it in the form of a reflective question. At first, we considered missing an abnormal result to be the outcome of regret (ie, "How would you feel if you missed an abnormal result?"). However, after discussion with adopters, we refined the anticipated regret content to more explicitly focus on the effect of missing an abnormal test result on patient outcomes (ie, "How would you feel if a patient had a poor outcome because you missed an abnormal test result?").

Tensions emerged when users' reactions conflicted with evidence about the effectiveness of the anticipated regret. Focus group participants expressed feelings of guilt and anger when they read the message and indicated that they did not want to have an emotional response when reading an email. Focus group participants described the anticipated regret content as "confrontational," "combative," and "not a nice way of starting."

**Table 1.** Participant characteristics for cocreation workshops.

Characteristic	All cocreation workshops (n=9)	Cocreation workshop #1 in Toronto, Ontario (n=7)	Cocreation workshop #2 in Kingston, Ontario (n=2)
<b>Gender, n</b>			
Male	2	1	1
Female	7	6	1
<b>Type of user, n</b>			
Adopter	9	7	2
Nonadopter	0	0	0
<b>Years practicing family medicine</b>			
Range	3-32	3-32	12-30
Average number of years	15	13	21

**Table 2.** Participant characteristics for focus groups.

Characteristic	Total for all focus groups (n=11)	Focus group #1 in Toronto, Ontario (n=3)	Focus group #2 in Kingston, Ontario (n=3)	Focus group #3 in Toronto, Ontario (n=5)
<b>Gender, n</b>				
Male	6	2	1	3
Female	5	1	2	2
<b>Type of user, n</b>				
Adopter	4	1	0	2
Nonadopter	7	2	3	3
<b>Years practicing family medicine</b>				
Range	4-32	4-27	18-30	5-32
Average number of years	20	19	24	19

One participant restated this message as:

*How would you feel if you were a murderer? How would you feel if you screwed up?* [SAR Non adopter P019]

Further, participants articulated that the content implied that they were not doing their job well. For example, one participant stated:

*You know it happens, but you don't have to sort of start off a positive report that says "we would really like...we think this report is really encouraging you to use it, it can help you with A, B, and C, but you know what really if you missed a test result you're shitty, you're really lousy, you're going to be sued, so use this tool."...that's just so condescending, negative on...you know the stress in our life is already there, we don't need to be told that we feel bad when things go bad.* [SAR Non adopter P016]

Exploration of reactions in focus groups revealed that the content, though evoking negative emotions, motivated several participants to read the email further; for example, one participant said:

*I would read more but I would be pissed off.* [SAR Non adopter P017]

Another participant acknowledged that the guilt was motivating, aligning with evidence behind the BCT:

*I think the first line is good, how would you feel if patient had a poor outcome because you missed an abnormal test result? I think that's every doctor's fear because we go through so many results and we're making decisions very quickly and it is possible to miss test results, which everybody's done. And sometimes you get lucky and it's not a big deal, and sometimes you don't and you feel horrible about it so I think the guilt thing in there is good, and it's motivating.* [SAR Non adopter P011]

Thus, although we recognized that too much fear can result in physicians avoiding the desired behavior (ie, logging in to the SAR), we believed that these strong negative reactions had the potential to engage recipients and maintain their attention, something that the previous SAR email was seemingly unable to do. We tried to avoid a judgmental tone that would evoke anger and to emphasize in the content that followed the anticipated regret message that logging in to the SAR could mitigate feelings of fear and anxiety about missing critical cancer screening results.

Participants also had negative reactions to the embedded problem-solving strategies for making time to access the SAR. The problem-solving BCT requires analysis of factors influencing the behavior and selecting strategies that overcome barriers or increase facilitators [11]. Our problem-solving strategies for improved SAR use were informed directly by adopters' real-world experiences, which included assigning a delegate to access the SAR on their behalf to overcome time pressures. One adopter described why nonadopters such as her colleagues need problem-solving strategies:

*Because it dumbs [accessing the SAR] down, it makes it a stepwise option for them to do. It's perfect. It's perfect. That's what they need. They just need someone to take them by the hand and say, 'Brian, print this out and give it to Dale and she will book you 15 minutes.' They can't find 15 minutes in their life to do this so this is a recipe for following through. It's really good. I like it a lot.* [SAR Adopter P008]

Nevertheless, several focus group participants expressed that they would not use these strategies, often stating that resource and time constraints prevented them from employing them in their practice. In some cases, participants disliked the strategies proposed because they felt that they were unnecessary; for example, one participant said that if they are going to do any clicking, they might as well click into the SAR rather than click into their calendar to make a reminder about logging into the SAR. Moreover, focus group participants described this strategy to book time in their calendar as "patronizing," "trivial," and "petty." For example, one participant stated:

*I think some of the tips are unnecessary. Like, book some calendar time. I know I can do that. I know it's one of those things everyone should fit in their schedule; however, they can't. So telling me to put it in your calendar, I'm a grown up.* [SAR Non adopter P018]

Though our goal was to provide strategies that helped users overcome time and workload challenges, participants often wanted to eliminate the BCT-informed content and recommended that we "just get to the bottom line- log in to the SAR."

### **Privacy Constraints Around Personalizing Unencrypted Emails With Performance Data**

One significant source of tension occurred when organizational privacy constraints restricted our ability to respond to some content and design changes offered by participants, especially in the case of personalizing content using individualized performance data and embedded performance comparator data.

When prompted with the following message "10,000 women in Ontario need follow-up to an abnormal pap test; is one of them yours?," focus group participants wanted the statement to be more personal and reflect the number of their patients that require follow-up:

*Send me an email saying I got 7 abnormal and give me a link to log in. You don't have to give me patient demographics and names or whatever. Or you can say, 'we've identified patients based on whatever you know, please log in to this to get your confidential list'...you're sending out an email that you want us to sort of attract physicians, you want to sort of catch them, you want to say. 'Hey listen, we've identified all these abnormal results, they've been sitting here in our whatever database.'...notify me that I have 8 patients, that I guarantee you would motivate me to open up and then follow-up but... [SAR Non adopter P019]*

*Yeah, the fact that I got an abnormal it's helpful for me to know. You're helping me because you're actually letting me know that there's some abnormal you want to make sure that I am aware of them and you want to make sure it's like you feel as if your part of a team, 'Hey we're sending you this as a reminder that you have some abnormal.' You sort of feel like it's not the stick, and the carrot is that you're actually improving on patient care. [SAR Non adopter P017]*

Though participants wanted varying degrees of personalized data (eg, Ontario rates vs individualized follow-up rates), it was clear that personalized data regarding the number of their own patients that required follow-up resonated with participants and had the potential to prompt SAR access.

In addition to personalized data, participants expressed a desire to have performance comparator data. Adopters described performance comparator information as “definitely the most unique thing” about the SAR in comparison to other tools used to monitor cancer screening in their practice, such as electronic medical records. One adopter expressed feeling comfort in numbers:

*Because you have no sense of [how] the rest of the family physicians in Ontario are functioning...you really do function in a bubble. [SAR Adopter P008]*

Similarly, focus group participants found this benefit of the SAR “very interesting” and noted that it sparked curiosity. One participant stated that they were interested in the comparison data from both a medical and legal standpoint:

*...I am interested in that if I have a glaring difference from other practices, I want to know that from both a medical and a legal point of view. [SAR Non adopter P013]*

Although most focus group participants agreed that comparison information was useful, they also reported that the content used to communicate this feature of the SAR was not as compelling as it could be. One participant suggested that the comparator data would be a strong motivator to use the SAR:

*If the data was planted right in the email. Like, this is you, this is people in the LHIN, and this is people in Ontario. [SAR Non adopter P012]*

Furthermore, focus group participants thought an image displaying their performance data would motivate them to log in to the SAR, especially if the image showed that they were underperforming in comparison to their peers:

*If there was an image there that I had a gauge of where I was so if I fell below the benchmark I'd probably feel pretty bad about it and I'd want to increase it. [SAR Non adopter P012]*

Our research team consulted the Cancer Care Ontario privacy team to understand which data level we could report in the email (eg, Ontario vs individual) and whether we could embed performance data in the content of the email. Performance data are considered personal information, and an error in sending could result in the disclosure of personal information of a physician to an unintended recipient. Given this potential privacy

breach, Cancer Care Ontario refrains from sharing any personal information in emails. Thus, content with personal data was not an option. However, reporting aggregate data at the Ontario, regional, or Local Health Integration Network (LHIN)-level was deemed appropriate because it could not be traced back to the performance of an individual physician. Our team decided that LHIN-level data would be the most compelling solution because it was the smallest unit of aggregate data that we could report. We attempted to make the concept about comparison data more salient by associating the use of the SAR for comparison data with social norms: “Thousands of Ontario family doctors access the SAR to compare their screening rates to other family doctors in Ontario and their region.”

### **Cocreation Methods Versus Study Objectives**

Our research team has extensive knowledge of behavior change theory and its application; however, at the beginning of this study, we had limited understanding of users' experience with the SAR and what email content would compel physicians to access the SAR. Accordingly, the research team made great efforts to create conditions for meaningful and substantial user participation in the development of emails. Adopter-generated emails developed during the first workshop yielded descriptions of the SAR and meaningful content about its benefits based on user experiences. However, most content did not explicitly align with a BCT and content that did was often limited to the material incentives BCT; for example, each adopter-generated email referenced how the SAR helps to track or achieve specified levels of preventive care needed to earn their annual Preventive Care Bonus. Moreover, participants often cited comparator information as a major benefit of the SAR, but we could not identify a BCT that accurately reflected this benefit. The research team encountered tensions around the desire to be true to cocreation methods that prioritize user-generated content and the need to accomplish the study objectives of defining, operationalizing, and testing distinct BCTs.

To address this tension, we used user-generated emails from the first workshop with adopters to act as a starting point for email design and made distinctions between “base content” and “variable content.” Base content was user-generated content that was iteratively refined throughout the process and would remain constant in all emails tested in the experiment; for example, content about how the SAR provides information about patients that have abnormal results and require follow-up was well-received by users and was considered critical information by the research team for all emails to be tested. Variable content, on the other hand, was user-informed BCT content that was created by the research team, focus-tested among participants, and tested in the factorial randomized experiment.

Iterative content development began with an open-ended cocreative exploration of effective communication about the SAR. During this stage, adopters played a significant role in generating content, especially non-BCT-informed content. However, as the content evolved, study and organizational constraints were sometimes in conflict with user feedback in our quest to have distinct and defined BCTs for testing within an applied setting. These constraints often took precedence,

transitioning the role of the participants from content generators to content informants as the project progressed. Further, the role of the research team moved from translators of user ideas and experiences to decision makers, ultimately, as illustrated in the tensions above. This transition of roles was necessary to accomplish the study's objectives of testing clearly operationalized BCTs.

## Discussion

### Principal Findings

This study involved the practical application of scientific evidence and methods to the development of emails to promote the use of an audit and feedback tool. We encountered 3 tensions that may be relevant for others who are considering cocreation methods to develop similar communication interventions.

The conflict between users' preferences and the broader scientific evidence regarding the potential effectiveness of a BCT highlights that what people seem to prefer and what works are not always the same. BCTs previously shown to be effective may not always elicit positive responses during cocreation or user testing. This tension is not unique to physician users because it also occurred in previous work with patient users relating to the development of mailings for people recovering from acute coronary syndrome [18]. Although users may be "experts of their experience" [19], they are not always experts in how to promote behavior change in themselves or in others [18]. It is necessary for researchers to be thoughtful when making design decisions because it may or may not be appropriate to make decisions solely based on whether user reactions to evidence-based content were positive or negative. It is the role of the researchers and designers to balance user feedback with the broader available evidence and explore the root cause of user reactions during data collection to make purposeful and informed design decisions.

Organizational constraints, including privacy policies, are a reality of health systems, which may limit the inclusion of personal performance data. Though users frequently recommended embedding personal performance data in the email, which was a finding consistent with best practices in audit and feedback design [20], privacy regulations did not permit our team to include such data in this setting. This created a situation in which users clearly indicated what they believed would help them and yet, we were unable to implement their suggestions.

Interestingly, the inclusion of individualized performance data may have expanded the scope of the email from an email intended to drive access to (and use of) an audit and feedback tool to an email that could actually be characterized as an audit and feedback intervention itself. There may be an opportunity in a different context with less stringent privacy regulations to provide some data, but not all data, to effectively bait the user to access the audit and feedback tool; for example, the email may present an overview of what kind of data the user could have access to if they were to use the tool. This may be more effective for communications with individuals who are underperformers rather than strong performers because

participants noted that data indicating underperformance would motivate action, and research has shown that feedback is most effective when baseline performance is low [21]. However, because we were not able to test this, we do not know if this strategy would, in fact, be effective at driving access to the SAR or other audit and feedback tools.

Cocreation challenges may have occurred throughout this study because our objective of developing an email with multiple BCTs required substantial knowledge of BCTs. Adopters were able to generate concepts and non-BCT-informed content in the early stages of the research, but their ability to participate in the cocreation of a product with defined research variables such as BCTs was, not surprisingly, limited thereafter. This tension between design goals to develop the best product, service, or tool for a given context and scientific goals to identify or test generalizable concepts or theory has been noted in other contexts [22]. Cocreation methods that invite users to engage in problem solving may be most appropriate in implementation research when the study (and design) objectives are flexible, fluid, and potentially user-driven; for example, cocreation methods may be helpful when engaging users to provide input on the design and functionality of products, such as an educational application aimed at improving knowledge of clinical skills among nursing students [23] or a complex health information system [24]. However, researchers and design teams are likely to face challenges using cocreation methods when products require the application of specific scientific knowledge and should consider the dynamic and changing role of the users from content generators to content informants as the product, service, or tool develops.

### Limitations

There are several limitations to this study. First, although we started with purposive sampling, we eventually turned to convenience sampling to recruit physicians from Toronto and Kingston. Convenience sampling could potentially contribute to selection bias and failure to recruit physicians with diverse views on how to communicate to physicians about accessing an audit and feedback tool. Furthermore, we did not purposively recruit based on adherence to cancer screening guidelines. Future research with this target audience may consider recruiting based on performance level to understand if reactions to BCT-informed content differ between high performers, average performers, and underperformers. Second, our findings occurred during the development of email communications about a specific audit and feedback tool created to help physicians monitor the cancer screening status of their rostered patients. These findings may or may not apply to the development of products on different communication channels, to the development of products that deal with a different audit and feedback tool, or to a different audience such as specialists rather than family physicians; for example, it may be possible that physicians could spontaneously generate BCT-informed content or products in other contexts. Third, the analysis of the 2×2×2 factorial experiment is currently underway, and we do not yet know the impact of the interventions developed.



## Conclusions

As audit and feedback tools proliferate across health care systems, there will be an increasing need for effective communications that promote the access and use of such tools. Teams within organizations that are tasked with developing content might consider cocreation methods because they allow for the development of communications that are both user- and BCT-informed. This paper describes a repeatable series of steps that teams could use when pursuing design of implementation interventions, beginning with engaging users to understand what

kind of content is compelling to them. Users provide honest feedback about what goes through their mind when they read the content and can provide thoughtful suggestions to inspire developers. However, teams working with users may need to consider whether and how they should balance user feedback with scientific evidence, organizational constraints, and study objectives. Thus, the role of users in decision making and that of BCTs along with the implementation context should all be considered during planning to determine whether cocreation methods would be appropriate to accomplish design objectives.

---

## Acknowledgments

This study was conducted with the support of the Ontario Institute for Cancer Research and Cancer Care Ontario through funding provided by the Government of Ontario. The research team would like to acknowledge the Primary Care, Privacy, and Operations teams at Cancer Care Ontario for their help in developing and reviewing content. More specifically, we would like to thank Matthew Jurczak for his diligent review of the content to ensure that it aligns with Cancer Care Ontario's privacy policies and Petra de Heer for her help to design and implement the emails.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Full list of behaviour change techniques considered during development process and definitions from Michie et al.'s behaviour change technique taxonomy v1[11].

[\[PDF File \(Adobe PDF File\), 208KB - humanfactors\\_v5i3e25\\_app1.pdf\]](#)

---

## Multimedia Appendix 2

Operationalizations of behaviour change techniques to be tested in the 2X2X2 factorial experiment and evidence to support testing.

[\[PDF File \(Adobe PDF File\), 329KB - humanfactors\\_v5i3e25\\_app2.pdf\]](#)

---

## Multimedia Appendix 3

Final email consisting of all BCT-informed content to be tested in the 2X2X2 factorial randomized experiment.

[\[PDF File \(Adobe PDF File\), 558KB - humanfactors\\_v5i3e25\\_app3.pdf\]](#)

---

## References

1. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci* 2006 Apr 28;1:9 [FREE Full text] [doi: [10.1186/1748-5908-1-9](https://doi.org/10.1186/1748-5908-1-9)] [Medline: [16722539](https://pubmed.ncbi.nlm.nih.gov/16722539/)]
2. Boyce MB, Browne JP, Greenhalgh J. Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: a qualitative study. *Implement Sci* 2014 Jun 27;9:84 [FREE Full text] [doi: [10.1186/1748-5908-9-84](https://doi.org/10.1186/1748-5908-9-84)] [Medline: [24972784](https://pubmed.ncbi.nlm.nih.gov/24972784/)]
3. Goyder C, Atherton H, Car M, Heneghan CJ, Car J. Email for clinical communication between healthcare professionals. *Cochrane Database Syst Rev* 2015;2:CD007979. [doi: [10.1002/14651858.CD007979.pub3](https://doi.org/10.1002/14651858.CD007979.pub3)] [Medline: [25698124](https://pubmed.ncbi.nlm.nih.gov/25698124/)]
4. Abdolrasulnia M, Collins BC, Casebeer L, Wall T, Spettell C, Ray MN, et al. Using email reminders to engage physicians in an Internet-based CME intervention. *BMC Med Educ* 2004 Sep 29;4:17 [FREE Full text] [doi: [10.1186/1472-6920-4-17](https://doi.org/10.1186/1472-6920-4-17)] [Medline: [15453911](https://pubmed.ncbi.nlm.nih.gov/15453911/)]
5. CCO, 2017. Reports Accessed by User Report. Unpublished internal dataset. CCO 2017 Jun 01.
6. CCO, 2017. SAR Refresh Notification MC Reports. Unpublished internal dataset. CCO 2017 May 01.
7. Badran H, Pluye P, Grad R. Advantages and disadvantages of educational email alerts for family physicians: viewpoint. *J Med Internet Res* 2015 Feb 27;17(2):e49 [FREE Full text] [doi: [10.2196/jmir.3773](https://doi.org/10.2196/jmir.3773)] [Medline: [25803184](https://pubmed.ncbi.nlm.nih.gov/25803184/)]
8. Vaisson G, Witteman HO, Bouck Z, Bravo CA, Desveaux L, Llovet D, et al. Testing behavior change techniques to encourage primary care physicians to access cancer screening audit and feedback reports: Protocol for a factorial randomized experiment of email content. *JMIR Res Protoc* 2018;7(1):e11. [doi: [10.2196/resprot.9090](https://doi.org/10.2196/resprot.9090)]

9. Gould NJ, Lorencatto F, Stanworth SJ, Michie S, Prior ME, Glidewell L, et al. Application of theory to enhance audit and feedback interventions to increase the uptake of evidence-based transfusion practice: an intervention development protocol. *Implement Sci* 2014 Jul 29;9:92 [FREE Full text] [doi: [10.1186/s13012-014-0092-1](https://doi.org/10.1186/s13012-014-0092-1)] [Medline: [25070404](https://pubmed.ncbi.nlm.nih.gov/25070404/)]
10. Smith WR. Evidence for the effectiveness of techniques To change physician behavior. *Chest* 2000 Aug;118(2 Suppl):8S-17S. [Medline: [10939994](https://pubmed.ncbi.nlm.nih.gov/10939994/)]
11. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013 Aug;46(1):81-95. [doi: [10.1007/s12160-013-9486-6](https://doi.org/10.1007/s12160-013-9486-6)] [Medline: [23512568](https://pubmed.ncbi.nlm.nih.gov/23512568/)]
12. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Health* 2015 Sep;42(5):533-544 [FREE Full text] [doi: [10.1007/s10488-013-0528-y](https://doi.org/10.1007/s10488-013-0528-y)] [Medline: [24193818](https://pubmed.ncbi.nlm.nih.gov/24193818/)]
13. Sedgwick P. Convenience sampling. *BMJ* 2013 Oct 25;347(oct25 2):f6304-f6304. [doi: [10.1136/bmj.f6304](https://doi.org/10.1136/bmj.f6304)]
14. Haggen P, Rowland N, Young G, McKemon S. Smallfire. A Framework for Thinking Through Co-design Workshops URL: <http://www.smallfire.co.nz/wp-content/uploads/2013/11/Co-designworkshopBusstopHandout.pdf> [accessed 2018-01-12] [WebCite Cache ID 6wQ90bc2d]
15. Djulbegovic B, Elqayam S, Reljic T, Hozo I, Miladinovic B, Tsalatsanis A, et al. How do physicians decide to treat: an empirical evaluation of the threshold model. *BMC Med Inform Decis Mak* 2014 Jun 05;14:47 [FREE Full text] [doi: [10.1186/1472-6947-14-47](https://doi.org/10.1186/1472-6947-14-47)] [Medline: [24903517](https://pubmed.ncbi.nlm.nih.gov/24903517/)]
16. Coricelli G. The potential role of regret in the physician-patient relationship: insights from neuroeconomics. *Adv Health Econ Health Serv Res* 2008;20:85-97. [Medline: [19552305](https://pubmed.ncbi.nlm.nih.gov/19552305/)]
17. Sorum PC, Mullet E, Shim J, Bonnin-Scaon S, Chasseigne G, Cogneau J. Avoidance of anticipated regret: the ordering of prostate-specific antigen tests. *Med Decis Making* 2004;24(2):149-159. [doi: [10.1177/0272989X04263163](https://doi.org/10.1177/0272989X04263163)] [Medline: [15090101](https://pubmed.ncbi.nlm.nih.gov/15090101/)]
18. Witteman HO, Presseau J, Nicholas AE, Jokhio I, Schwalm JD, Grimshaw JM, et al. Negotiating Tensions Between Theory and Design in the Development of Mailings for People Recovering From Acute Coronary Syndrome. *JMIR Hum Factors* 2017 Mar 01;4(1):e6 [FREE Full text] [doi: [10.2196/humanfactors.6502](https://doi.org/10.2196/humanfactors.6502)] [Medline: [28249831](https://pubmed.ncbi.nlm.nih.gov/28249831/)]
19. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. *CoDesign* 2008 Jun 24;4(1):5-18. [doi: [10.1080/15710880701875068](https://doi.org/10.1080/15710880701875068)]
20. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci* 2014 Jan 17;9:14 [FREE Full text] [doi: [10.1186/1748-5908-9-14](https://doi.org/10.1186/1748-5908-9-14)] [Medline: [24438584](https://pubmed.ncbi.nlm.nih.gov/24438584/)]
21. Ivers NM, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012 Jun 13(6):CD000259. [doi: [10.1002/14651858.CD000259.pub3](https://doi.org/10.1002/14651858.CD000259.pub3)] [Medline: [22696318](https://pubmed.ncbi.nlm.nih.gov/22696318/)]
22. Grocott L, Kobori M. The Affordances of Designing for the Learning Sciences. : Aalto University; 2015 Jun 25 Presented at: Proceedings of the 3rd International Conference for Design Education Researchers; 2015 June 3; Chicago, USA p. 1180-1195 URL: <https://tinyurl.com/y999pxup> [doi: [10.13140/RG.2.1.2904.6880](https://doi.org/10.13140/RG.2.1.2904.6880)]
23. O'Connor S, Andrews T. Using Co-Design with Nursing Students to Create Educational Apps for Clinical Training. *Stud Health Technol Inform* 2016;225:334-338. [Medline: [27332217](https://pubmed.ncbi.nlm.nih.gov/27332217/)]
24. Kushniruk A, Nørh C. Participatory Design, User Involvement and Health IT Evaluation. *Stud Health Technol Inform* 2016;222:139-151. [Medline: [27198099](https://pubmed.ncbi.nlm.nih.gov/27198099/)]
25. St QT, Brunton JA. Implicit Processes, Self-Regulation, and Interventions for Behavior Change. *Front Psychol* 2017;8:346 [FREE Full text] [doi: [10.3389/fpsyg.2017.00346](https://doi.org/10.3389/fpsyg.2017.00346)] [Medline: [28337164](https://pubmed.ncbi.nlm.nih.gov/28337164/)]
26. Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing physician behavior: what works? *Am J Manag Care* 2015 Jan;21(1):75-84 [FREE Full text] [Medline: [25880152](https://pubmed.ncbi.nlm.nih.gov/25880152/)]
27. Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database Syst Rev* 2011 Jul 06(7):CD009255 [FREE Full text] [doi: [10.1002/14651858.CD009255](https://doi.org/10.1002/14651858.CD009255)] [Medline: [21735443](https://pubmed.ncbi.nlm.nih.gov/21735443/)]
28. Kiran T, Wilton AS, Moineddin R, Paszat L, Glazier RH. Effect of payment incentives on cancer screening in Ontario primary care. *Ann Fam Med* 2014 Jul;12(4):317-323 [FREE Full text] [doi: [10.1370/afm.1664](https://doi.org/10.1370/afm.1664)] [Medline: [25024239](https://pubmed.ncbi.nlm.nih.gov/25024239/)]
29. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. *Health Psychol Rev* 2015 Sep 16:1-24. [doi: [10.1080/17437199.2015.1082146](https://doi.org/10.1080/17437199.2015.1082146)] [Medline: [26315814](https://pubmed.ncbi.nlm.nih.gov/26315814/)]
30. Kozica S, Lombard C, Teede H, Ilic D, Murphy K, Harrison C. Initiating and continuing behaviour change within a weight gain prevention trial: a qualitative investigation. *PLoS One* 2015;10(4):e0119773 [FREE Full text] [doi: [10.1371/journal.pone.0119773](https://doi.org/10.1371/journal.pone.0119773)] [Medline: [25875943](https://pubmed.ncbi.nlm.nih.gov/25875943/)]
31. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. *Int J Behav Nutr Phys Act* 2017 Dec 28;14(1):42 [FREE Full text] [doi: [10.1186/s12966-017-0494-y](https://doi.org/10.1186/s12966-017-0494-y)] [Medline: [28351367](https://pubmed.ncbi.nlm.nih.gov/28351367/)]

---

**Abbreviations**

**BCT:** behavior change technique

**LHIN:** local health integration network

**SAR:** Screening Activity Report

---

*Edited by G Eysenbach; submitted 18.01.18; peer-reviewed by TH van de Belt, T Risling, Y van Kasteren; comments to author 15.02.18; revised version received 25.05.18; accepted 18.06.18; published 04.09.18.*

*Please cite as:*

*Bravo CA, Llovet D, Witteman HO, Desveaux L, Presseau J, Saragosa M, Vaisson G, Umar S, Timmouth J, Ivers NM*

*Designing Emails Aimed at Increasing Family Physicians' Use of a Web-Based Audit and Feedback Tool to Improve Cancer Screening Rates: Cocreation Process*

*JMIR Hum Factors 2018;5(3):e25*

*URL: <http://humanfactors.jmir.org/2018/3/e25/>*

*doi: [10.2196/humanfactors.9875](https://doi.org/10.2196/humanfactors.9875)*

*PMID: [30181108](https://pubmed.ncbi.nlm.nih.gov/30181108/)*

©Caroline A. Bravo, Diego Llovet, Holly O Witteman, Laura Desveaux, Justin Presseau, Marianne Saragosa, Gratianna Vaisson, Shama Umar, Jill Timmouth, Noah M. Ivers. Originally published in JMIR Human Factors (<http://humanfactors.jmir.org>), 04.09.2018. 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 <http://humanfactors.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# A Patient-Facing Diabetes Dashboard Embedded in a Patient Web Portal: Design Sprint and Usability Testing

William Martinez<sup>1</sup>, MD, MS; Anthony L Threatt<sup>2</sup>, MArch, PhD; S Trent Rosenbloom<sup>3,4,5</sup>, MD, MPH; Kenneth A Wallston<sup>6</sup>, PhD; Gerald B Hickson<sup>7</sup>, MD; Tom A Elasy<sup>1</sup>, MD, MPH

<sup>1</sup>Division of General Internal Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>2</sup>Health Information Technology, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>3</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>4</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>5</sup>Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>6</sup>School of Nursing, Vanderbilt University, Nashville, TN, United States

<sup>7</sup>Quality, Safety & Risk Prevention, Vanderbilt University Medical Center, Nashville, TN, United States

**Corresponding Author:**

William Martinez, MD, MS

Division of General Internal Medicine and Public Health

Vanderbilt University Medical Center

2525 West End Avenue

Suite 450

Nashville, TN, 37203

United States

Phone: 1 615 322 7277

Fax: 1 615 936 1269

Email: [william.martinez@vumc.org](mailto:william.martinez@vumc.org)

## Abstract

**Background:** Health apps and Web-based interventions designed for patients with diabetes offer novel and scalable approaches to engage patients and improve outcomes. However, careful attention to the design and usability of these apps and Web-based interventions is essential to reduce the barriers to engagement and maximize use.

**Objective:** The aim of this study was to apply design sprint methodology paired with mixed-methods, task-based usability testing to design and evaluate an innovative, patient-facing diabetes dashboard embedded in an existing patient portal and integrated into an electronic health record.

**Methods:** We applied a 5-day design sprint methodology developed by Google Ventures (Alphabet Inc, Mountain View, CA) to create our initial dashboard prototype. We identified recommended strategies from the literature for using patient-facing technologies to enhance patient activation and designed a dashboard functionality to match each strategy. We then conducted a mixed-methods, task-based usability assessment of dashboard prototypes with individual patients. Measures included validated metrics of task performance on 5 common and standardized tasks, semistructured interviews, and a validated usability satisfaction questionnaire. After each round of usability testing, we revised the dashboard prototype in response to usability findings before the next round of testing until the majority of participants successfully completed tasks, expressed high satisfaction, and identified no new usability concerns (ie, stop criterion was met).

**Results:** The sample (N=14) comprised 5 patients in round 1, 3 patients in round 2, and 6 patients in round 3, at which point we reached our stop criterion. The participants' mean age was 63 years (range 45-78 years), 57% (8/14) were female, and 50% (7/14) were white. Our design sprint yielded an initial patient-facing diabetes dashboard prototype that displayed and summarized 5 measures of patients' diabetes health status (eg, hemoglobin A<sub>1c</sub>). The dashboard used graphics to visualize and summarize health data and reinforce understanding, incorporated motivational strategies (eg, social comparisons and gamification), and provided educational resources and secure-messaging capability. More than 80% of participants were able to successfully complete all 5 tasks using the final prototype. Interviews revealed usability concerns with design, the efficiency of use, and content and terminology, which led to improvements. Overall satisfaction (0=worst and 7=best) improved from the initial to the final prototype (mean 5.8, SD 0.4 vs mean 6.7, SD 0.5).

**Conclusions:** Our results demonstrate the utility of the design sprint methodology paired with mixed-methods, task-based usability testing to efficiently and effectively design a patient-facing, Web-based diabetes dashboard that is satisfying for patients to use.

(*JMIR Hum Factors* 2018;5(3):e26) doi:[10.2196/humanfactors.9569](https://doi.org/10.2196/humanfactors.9569)

## KEYWORDS

diabetes mellitus, type 2; patient portals; qualitative research; consumer health informatics

## Introduction

### Background

Diabetes is a leading cause of kidney failure, heart disease, stroke, visual impairment, and nontraumatic lower limb amputations [1]. Many of these complications can be delayed or prevented through disease control. Research demonstrates that diabetes self-monitoring, preventative health services, medication adherence, regular exercise, and attention to diet can lead to improved outcomes [2,3]. Despite their importance, few patients consistently receive all recommended services or engage in recommended self-care behaviors that can be challenging to implement and sustain [4,5]. Many patients with diabetes struggle with the knowledge and motivation necessary to successfully manage their disease [6].

Interventions aimed at enhancing patients' motivation, skills, knowledge, and confidence in diabetes self-care have had limited success, with many relying on face-to-face interactions that are costly and challenging to scale [7,8]. Web-based diabetes self-management interventions have the potential to overcome these limitations; however, these interventions have also demonstrated variable effects on patients' self-care and glycemic control [9,10]. Mixed results have been attributed to differences in the design and usability of these Web-based interventions, leading to varying degrees of user engagement [10,11]. Web-based interventions with greater user engagement are associated with better outcomes [12,13]. However, some Web-based interventions have not involved end users in the design process [14,15], and many have failed to include one or more recommended features for increasing patient engagement, including (1) ability to track, visualize, and summarize health data; (2) guidance in response to the data displayed; (3) ability to communicate with health care providers; (4) peer support; and (5) motivational challenges using elements of game design and competition [11,16].

Human-centered design is an approach to software development that emphasizes optimal user experience by integrating users directly into the design process and helps ensure the creation of a suitable user interface [17,18]. One human-centered design method, called design sprint, is a rapid 5-phase user-centered process that utilizes design principles to understand the problem, explore creative solutions, identify and map the best ideas, prototype, and ultimately test [17,18]. Usability testing ensures that Web-based interventions meet users' expectations and work as intended, such that users are able to efficiently and effectively interact with the website [11]. Although usability testing is sometimes performed once the Web-based intervention has been fully developed, incorporating usability testing into the design

process beginning with the earliest prototype provides the greatest opportunity to inform and improve the user interface design [17,18].

### Objectives

This paper describes the application of design sprint methodology paired with mixed-methods, task-based usability testing to design and evaluate an innovative, patient-facing diabetes dashboard embedded in an existing patient portal, My Health at Vanderbilt (MHAV) [19] and integrated into an electronic health record. In particular, we sought to design a dashboard that addresses the needs of users, allows users to easily comprehend their diabetes health data, incorporates recommended strategies for increasing user engagement, and is satisfying and easy to use.

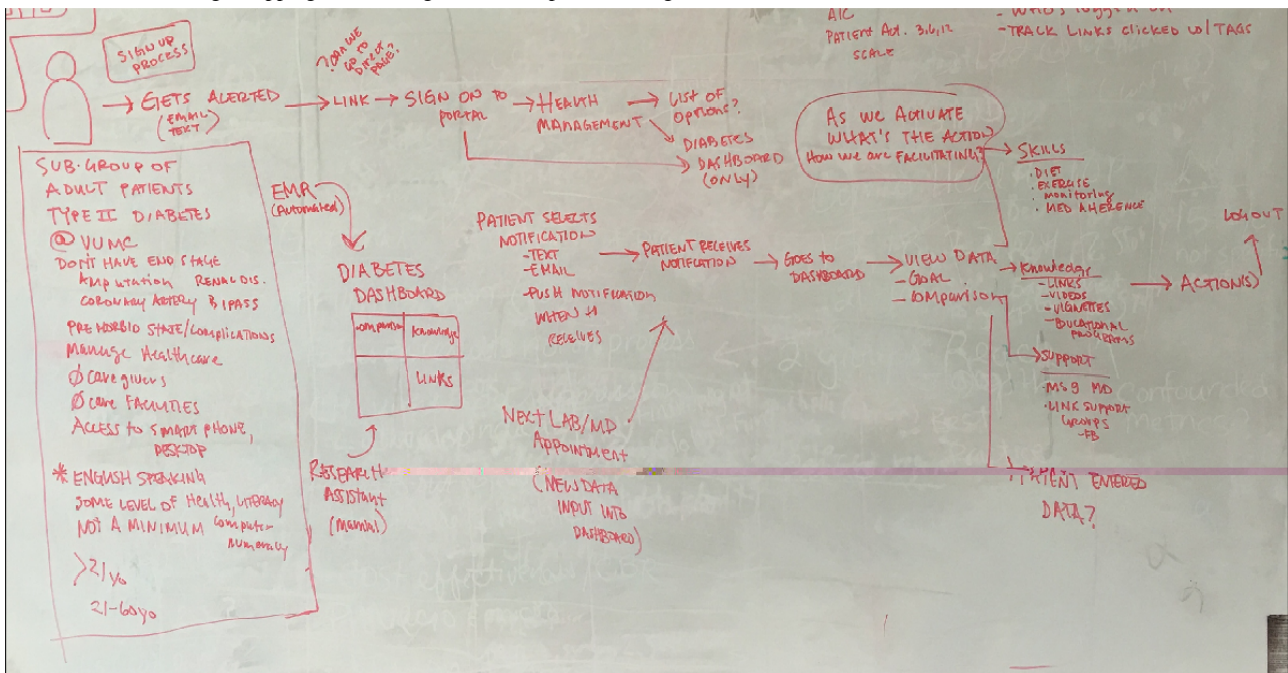
## Methods

### Dashboard Design

We utilized a 5-day design sprint methodology [17,18] developed by Google Ventures (Alphabet Inc, Mountain View, CA) to design our initial dashboard prototype. The process was facilitated by an experienced health information technology expert (ALT) who specializes in user experience (UX) and product design. A 5-day design sprint approach was chosen over other iterative agile methodologies because a design sprint approach offered the ability to rapidly develop a user-centered solution in the form of a prototype that could be tested and revised before investing limited research funds into the programming of the dashboard.

On day 1, we began by mapping out our challenge (Figure 1) to create a dashboard that would satisfy patients' desire for information regarding their diabetes health status and address existing challenges in patients' diabetes knowledge and motivation for diabetes self-management [5,20]. This process was informed by a review of the literature [14,21-30] from which we identified factors contributing to the limited efficacy of existing digital interventions, including (1) absence of user-centered design [14], (2) lack of integration with the health care delivery system [22,28], (3) absence of key features to maximize patient engagement, including patient-centered motivational strategies [29], and (4) failure to account for the unique needs of older patients and those with limited health literacy [30-32]. In addition, we reviewed recommended strategies to increase patient activation [6,33] (ie, the motivation, knowledge, skills, and confidence for managing one's health condition) using mobile apps [16] and prior research on the potential role of social comparison information for motivating diabetes self-care [27,34].

**Figure 1.** Whiteboard image mapping out challenge to create a patient-facing, diabetes dashboard.



We also met one-on-one with expert stakeholders (eg, patient portal users with diabetes, diabetes educators, behavioral scientists, physicians, educators, and nurses) to ask questions aimed at enhancing our understanding of the challenge and refine our map. We identified expert stakeholders by approaching organizational leaders with a description of the project and by asking them to identify individuals in their area who could provide valuable input. For example, we approached the director of the Vanderbilt University Hospital Patient and Family Advisory Council who connected us with patients from the Council, who had diabetes, were current patient portal users, and expressed interest in improving care for people with diabetes. Experts' comments were recorded in the form of *how might we* (HMW) statements [17,18]. The HMW method is used in design thinking to take insights and challenges and reframe them as opportunities [17,18]. Consistent with design sprint methodology, experts' HMW statements were reviewed (Figure 2) to identify statements that shared a common theme. This was followed by grouping the statements into categories based on emerging themes to identify the most useful ideas for building the prototype. Experts encouraged the authors to consider how we might design the dashboard to (1) maximize accessibility, (2) frame diabetes health data in ways that promote patients' understanding and motivate health behaviors, (3) facilitate patient action in response to the data they see (eg, patient resources and referral services), (4) enable communication with their health care team, (5) enhance social supports, and (6) incorporate strategies (eg, goal setting, progress tracking, and positive reinforcement) that motivate health behavior and keep users engaged.

On day 2, the existing ideas, architecture, and designs from health care and other industries related to the challenge were reviewed to establish the building blocks of our prototype. For example, existing solutions for displaying health and performance data and other types of quantitative, longitudinal, and benchmarked data from other industries (eg, finance and

education) were reviewed. Subsequently, findings from the review and the meetings with expert stakeholders were used to sketch our own solutions (Figure 3).

On day 3, the solutions were critiqued and the solutions that had the greatest potential to successfully meet the challenge in the long term were decided by consensus. Following this, the authors adapted the solutions chosen to create a storyboard or step-by-step plan for the prototype (Figure 4).

On day 4, the authors developed the prototype using Apple Keynote (Apple Inc, Cupertino, CA) [35]. They collected assets (eg, stock imagery or icons) and stitched all components of the prototype together. Keynote slides (ie, screens) were tethered together using the *animate* feature to transition from one slide (ie, screen) to the next based on the action the user performs within the prototype. This resulted in an initial prototype (Figure 5) that functioned similar to a real webpage and was ready for the first round of usability testing on day 5. The initial prototype displayed and summarized 5 measures of patients' diabetes health status (ie, hemoglobin A<sub>1c</sub> [HbA<sub>1c</sub>], systolic blood pressure, low-density lipoprotein cholesterol, microalbumin, and flu vaccination status). The existing literature on patient's information needs when interpreting test results and strategies for improving comprehension was reviewed [36-38]. In addition, the authors identified recommended strategies for using patient-facing technologies to increase patient activation and incorporated dashboard functionality that matched each strategy. For example, for each measure, the dashboard used graphics to visualize and summarize health data and reinforce understanding with a color-coded system (red, yellow, and green) similar to the National Heart, Lung, and Blood Institute's asthma treatment guideline [39] to indicate when action is needed. To facilitate understanding, we paired each measure with hyperlinks to literacy level-appropriate educational materials. To help motivate patients, the dashboard provided patients with social and goal-based comparison information regarding their diabetes

health status [27,34]. In addition, using elements of game design, a star rating provided patients with feedback on the number of measures at goal. To facilitate communication with their health care team, patients could click a link to contact their doctor's office via a secure message. Reminders for self-care (eg, take medication, exercise, etc) could be set and delivered to patients' mobile phones or email, and diabetes self-care goals could be set and tracked.

prototypes with individual patients under controlled conditions. Patients were recruited from the Vanderbilt Adult Primary Care (VAPC) clinic. Individual usability sessions lasted between 30 and 75 min. Given that the majority of usability problems are commonly identified within the first 5 usability evaluations [40-42], each round of usability testing included between 3 and 6 participants. After each round of usability testing, the dashboard prototype was revised in response to usability findings before the next round of testing.

Usability Study Design

From September to October 2016, we conducted a mixed-methods, task-based usability study of dashboard

Figure 2. Design sprint day 1—expert comments/ideas organized into categories.

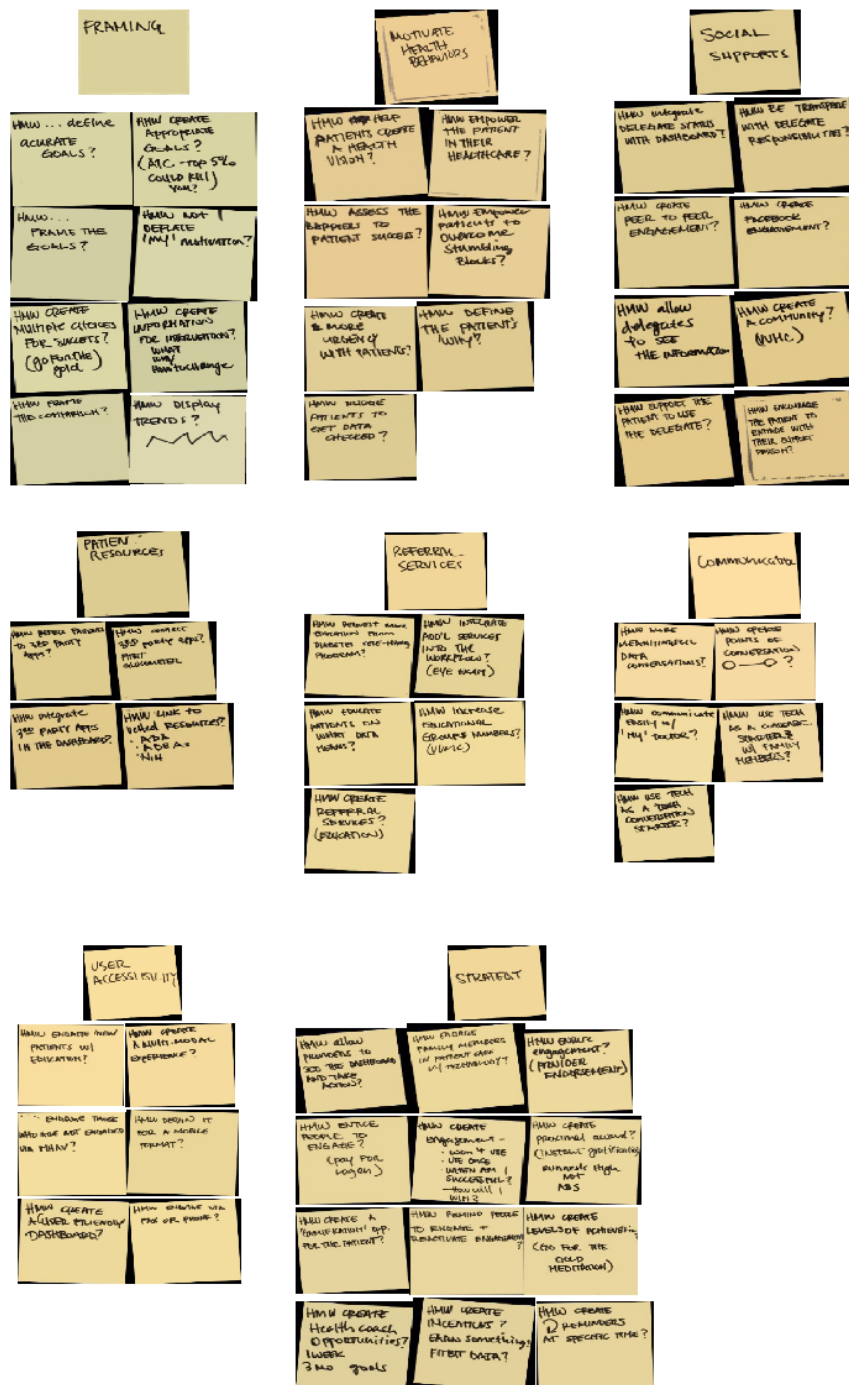


Figure 3. Design sprint day 2—solution sketches.

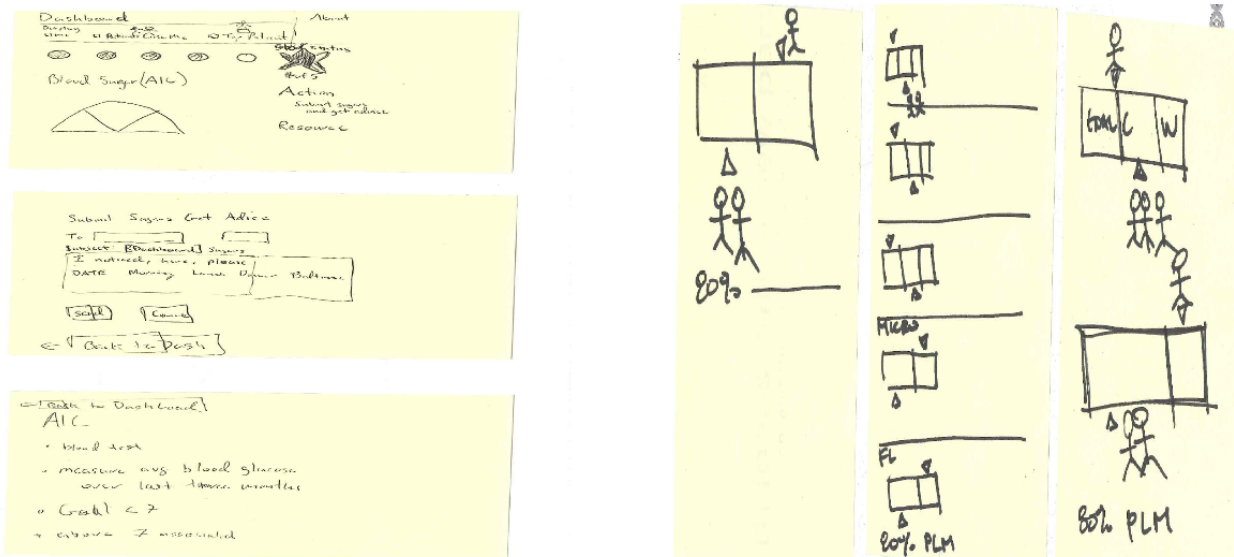


Figure 4. Design sprint day 3—dashboard storyboard.

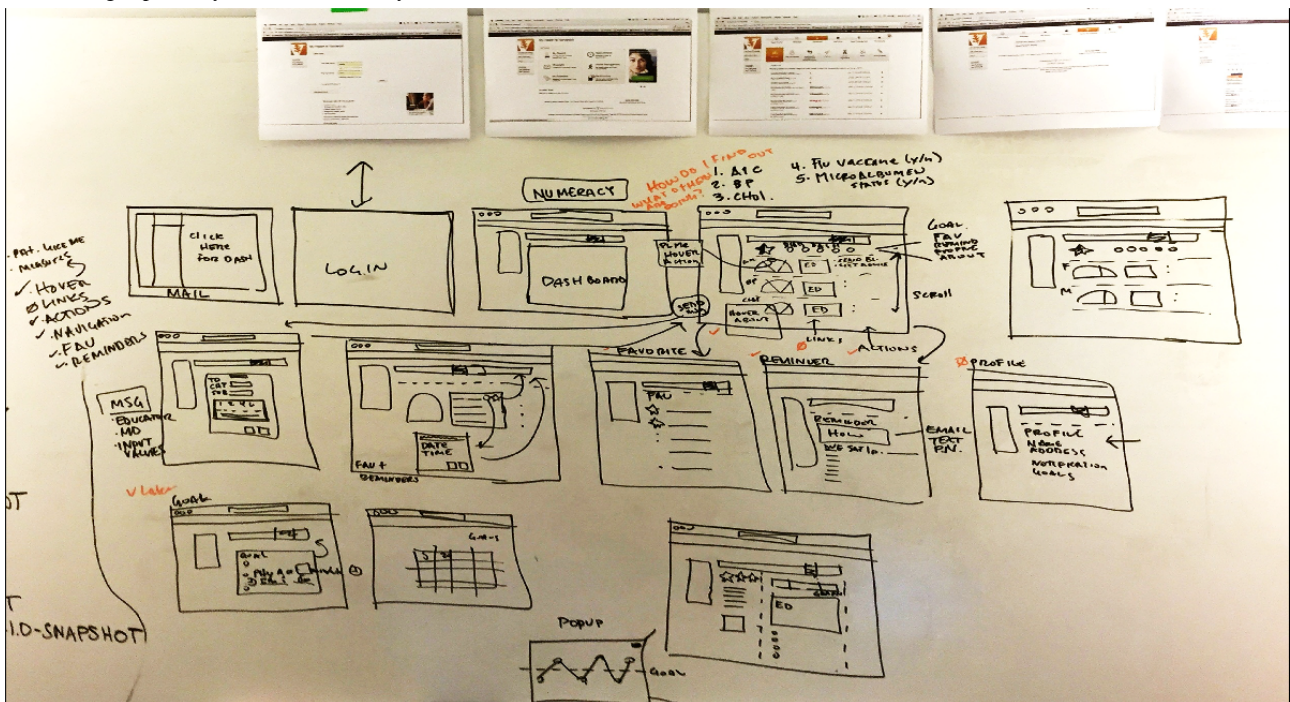
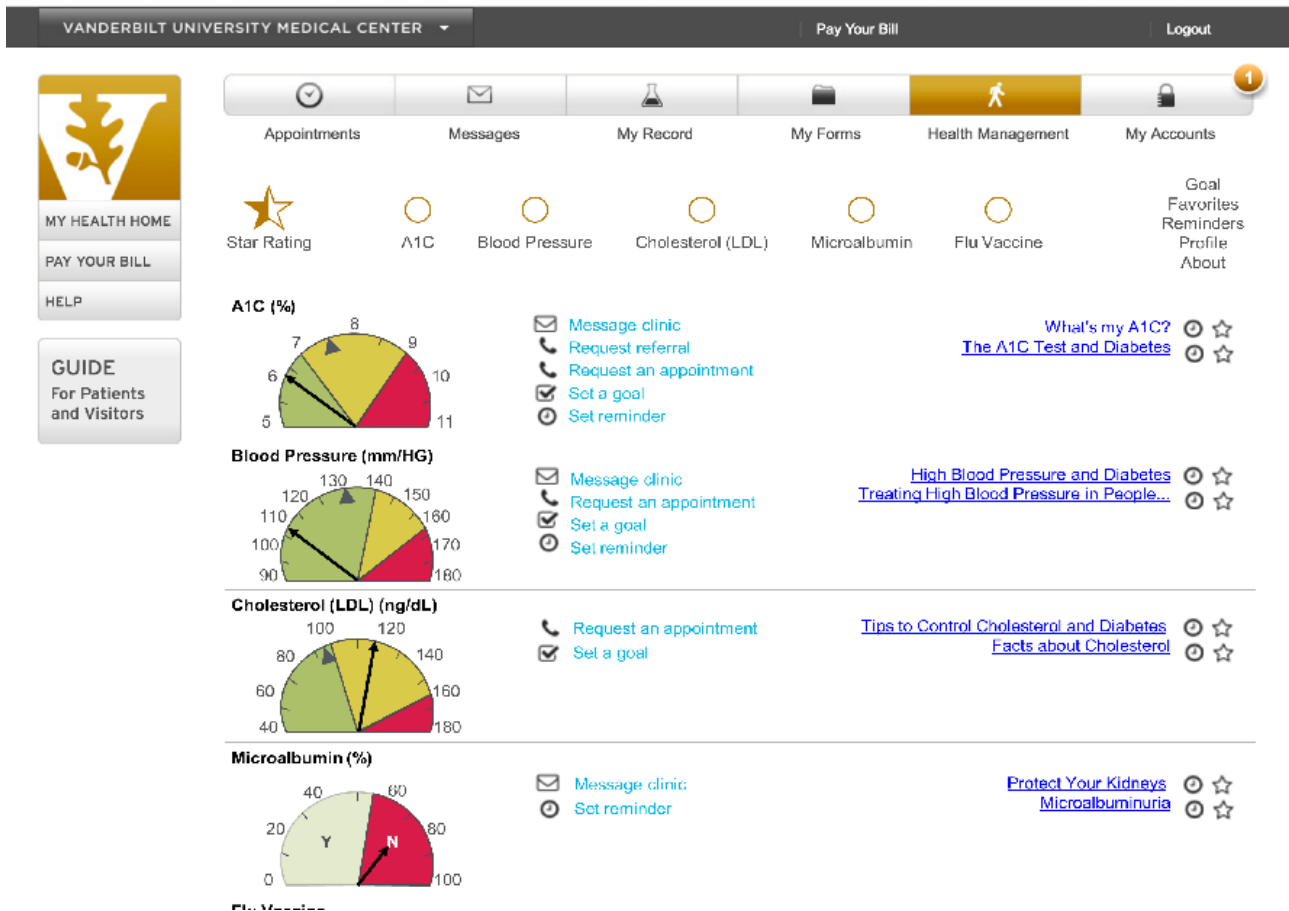




Figure 5. Design sprint day 4—screenshot of initial dashboard prototype. A<sub>1c</sub>: hemoglobin A<sub>1c</sub>.



**Setting**

The VAPC clinic is located within the Vanderbilt University Medical Center (VUMC) in Nashville, TN. The clinic cares for about 25,000 unique patients annually, of which about 4500 (18.00%) have diabetes. All clinical data are entered into an electronic health record, and the patients are provided access to their clinical data via a Web portal.

**Participants and Recruitment**

Participants were eligible for the study if they had type 2 diabetes mellitus, were English-speaking, were aged 21 years or older, and were current users of the VUMC patient Web portal, MHAV. Potential participants were identified automatically using VUMC’s Subject Locator to query the electronic health records of patients with upcoming clinic appointments for discrete inclusion and exclusion criteria. Identified patients (n=334) were mailed a letter describing the study and asked to contact the investigators if they were interested in participating. Interested patients (n=22) contacted the research coordinator to learn more about the study and confirm eligibility. Patients who agreed to participate (n=17) were scheduled to participate in a usability session on the day of their clinic appointment. Overall, 3 patients canceled due to weather or a conflicting appointment. A total of 14 patients ultimately completed a usability session and provided written informed consent before participating in their session. The Vanderbilt University Institutional Review Board approved this research.

**Data Collection and Measures**

Before the usability testing session, enrolled patients were asked to complete a short questionnaire before their interview. The questionnaire included basic demographic questions, including items about computer and smartphone usage and internet access, as well as validated measures of health literacy [43] and numeracy [44]. In addition, data regarding comorbidities were extracted from participants’ medical record as reported by the physicians within the patients’ problem list.

Each participant received a standardized introduction to the dashboard and the *think-aloud* procedure that allows testing observers to understand and track a participant’s thought processes as they navigate the dashboard [45]. One of the authors (ALT) led each session using a semistructured interview guide, while another author (WM) observed and took notes. With a dashboard prototype that contained fictitious patient data, participants were asked to perform common standardized tasks including logging in, retrieving HbA<sub>1c</sub> data, messaging their doctor, setting a reminder, and setting a goal. The tasks were designed to represent what typical users might do when visiting their dashboard. All participants accessed and navigated the dashboard using a 15-inch MacBook Pro 11,3 (2014 generation) with an external mouse and Chrome Web browser with default resolution. In addition, after participants attempted each assigned task (eg, message your doctor), the interviewers used open-ended questions outlined in the interview guide to elicit participants’ (1) expectations for the feature’s functionality, (2) ability to comprehend the information

displayed, (3) ability to navigate to and from the feature, (4) satisfaction with the feature, and (5) how the feature might be improved. Each session was audio-recorded, and the computer screen was video-recorded using QuickTime Player (Apple Inc, Cupertino, CA).

To assess and quantify participant satisfaction with the dashboard, at the conclusion of their usability session, participants completed 12 items from the Computer System Usability Questionnaire (CSUQ), which assess participants' perceptions of the dashboard's ease of use, likability of the interface, and overall satisfaction using a 7-point Likert response scale (1=strongly disagree to 7=strongly agree), with 7 indicating the highest possible satisfaction [46].

## Data Analysis

### Task Completion Analysis

Task completion was coded with a usability rating scale utilized in prior studies [47-49]. Task completion was rated on a 5-category scale: (1) successful/straightforward, (2) successful/prolonged, (3) partial, (4) unsuccessful/prolonged, and (5) gave up [47]. Two coders first coded the same usability session video (not used in the analysis) to calibrate their coding. They subsequently coded the remaining videos independently. Disagreements were resolved by consensus, and both coders were blinded to the dashboard prototype representing the initial prototype and the prototypes that were revisions.

### Interview Analysis

Audio files of interviews were submitted to a professional transcription service, Rev.com Inc (San Francisco, CA). Transcripts were checked for accuracy and identifying information was removed. Deidentified transcripts were imported into NVivo 10 (version 10; QSR International, Burlington, VT) for coding and analysis. Similar to other health app usability studies [47,50], we used selective coding to capture participants' comments about usability concerns [51]. Participant comments were sorted into categories that addressed 3 elements of usability: design, efficiency of use, and content and terminology [52]. A research assistant with training in qualitative methods coded all interviews. After the initial coding, a second trained coder reviewed each code and noted any discrepancies. The 2 coders then met and resolved any differences by consensus. Illustrative quotes from participants were edited slightly for grammar and clarity for inclusion in this paper. Participants' comments informed revisions to the dashboard prototype.

### Statistical Analysis

Descriptive statistics were used to characterize the study participants, task completion, and survey data. All analyses were completed with SAS version 9.4 (SAS Institute, Inc, Cary, NC).

### Stop Criteria

Data analysis began after the initial round of testing, and the authors used the findings to inform prototype revisions before the subsequent round of testing. Additional rounds of testing were conducted until the majority of participants within a round of testing (1) were able to successfully complete all tasks, (2)

indicated high overall satisfaction with the dashboard as assessed by the overall satisfaction item on the CSUQ (score $\geq$ 6), and (3) expressed no new usability concerns during the interview (ie, saturation).

## Results

### Participants

Table 1 shows participant characteristics. The sample (N=14) comprised 5 patients in round 1, 3 patients in round 2, and 6 patients in round 3; at this point, the authors reached their stop criteria. Participants' mean age was 63 years (range 45-78 years), 57% (8/14) were female, and 50% (7/14) were white. All participants reported using a home computer, and 64% (9/14) reported using a smartphone. All participants had home internet access. Most participants had one or more comorbid diseases in addition to diabetes.

### Task-Based Usability

Figure 6 illustrates task performance among the 5 participants in round 1 who tested the initial prototype compared with the 6 participants in round 3 who tested the final prototype. Participants attempted 5 tasks that ranged in complexity from logging in to setting a reminder.

#### Tasks: (A) Log-In and (B) Set a Goal

All participants in both rounds straightforwardly logged in to the dashboard and set a goal.

#### Task: (C) Identify Most Recent Hemoglobin A<sub>1c</sub>

Only one participant in the initial round of testing was able to identify their most recent HbA<sub>1c</sub> value from the dashboard. Most participants had difficulty interpreting the dial display, were confused regarding which icon on the dial indicated the user's most current value, and could not comprehend the HbA<sub>1c</sub> data. In response, the authors revised the data display design and status indicator icons. They relocated the features aimed at facilitating patients' understanding of their health data, including a hover over info icon providing a nontechnical description of the measure (eg, HbA<sub>1c</sub>) and links to literacy level-sensitive educational materials so they were adjacent to the data (see Figure 1 initial prototype and Figure 7 final prototype). After revisions, all 6 participants in the final round were able to complete the task and comprehend their data.

#### Task: (D) Message Doctor's Office

All 5 participants in the initial round were able to message their doctor's office; however, 2 participants hesitated or demonstrated some confusion despite completing the task. Participants indicated that they were accustomed to using the existing messaging icon within the header of the patient portal, and some struggled to locate the messaging icon within the dashboard. After revising the icon in response to feedback (ie, larger text, adding color and a button icon), the majority of participants in the final round successfully completed the task. However, 3 participants continued to initially attempt messaging via the existing icon in the header, one of whom completed the task only after being directed to the correct button icon.

**Table 1.** Participant characteristics.

Characteristic	Total (N=14)	Round 1 (N=5)	Round 2 (N=3)	Round 3 (N=6)
Age (years), mean (SD)	63.4 (11.0)	62.2 (10.3)	75.7 (3.2)	58.2 (9.9)
<b>Age (years), n (%)</b>				
40-49	1 (7)	0 (0)	0 (0)	1 (16)
50-59	4 (29)	2 (40)	0 (0)	2 (33)
60-69	4 (29)	2 (40)	0 (0)	2 (33)
70-79	5 (36)	1 (20)	3 (100)	1 (16)
<b>Gender, n (%)</b>				
Female	8 (57)	3 (60)	0 (0)	5 (83)
Male	6 (43)	2 (40)	3 (100)	1 (17)
<b>Race, n (%)</b>				
White	7 (50)	3 (60)	1 (33)	3 (50)
African American	3 (21)	1 (20)	1 (33)	1 (17)
Asian	2 (14)	1 (20)	1 (33)	0 (0)
Other	2 (14)	0 (0)	0 (0)	2 (33)
<b>Education, n (%)</b>				
High school degree / graduate equivalency degree	1 (7)	1 (20)	0 (0)	0 (0)
Some college	3 (21)	1 (20)	0 (0)	2 (33)
College degree	5 (36)	1 (20)	2 (67)	2 (33)
Postgraduate degree	5 (36)	2 (40)	1 (33)	2 (33)
Health literacy, mean (range <sup>a</sup> )	13.4 (11-15)	13.2 (12-15)	12.7 (11-15)	14.0 (13-15)
Numeracy, mean (range <sup>b</sup> )	15.0 (7-18)	13.0 (7-18)	17.0 (16-18)	15.7 (10-18)
Home computer user <sup>c</sup> , n (%)	14 (100)	5 (100)	3 (100)	6 (100)
Smartphone user, n (%)	9 (64)	3 (60)	2 (67)	4 (67)
Home internet access, n (%)	14 (100)	5 (100)	3 (100)	6 (100)
<b>Comorbidities, n (%)</b>				
Hyperlipidemia	10 (71)	3 (60)	3 (100)	4 (67)
Atherosclerotic cardiovascular disease	3 (21)	0 (0)	1 (33)	2 (33)
Hypertension	7 (50)	2 (40)	3 (100)	2 (33)
Chronic kidney disease	3 (21)	1 (20)	1 (33)	1 (17)

<sup>a</sup>Possible score range: 3 (worst) to 15 (best).

<sup>b</sup>Possible score range: 3 (worst) to 18 (best).

<sup>c</sup>Includes desktops, laptops, or tablets.

**Figure 6.** Task-based usability ratings for initial and final prototype iterations. The asterisk indicates that one participant within the final round of testing was not asked to complete the task due to time constraints. HbA1c: hemoglobin A1c.

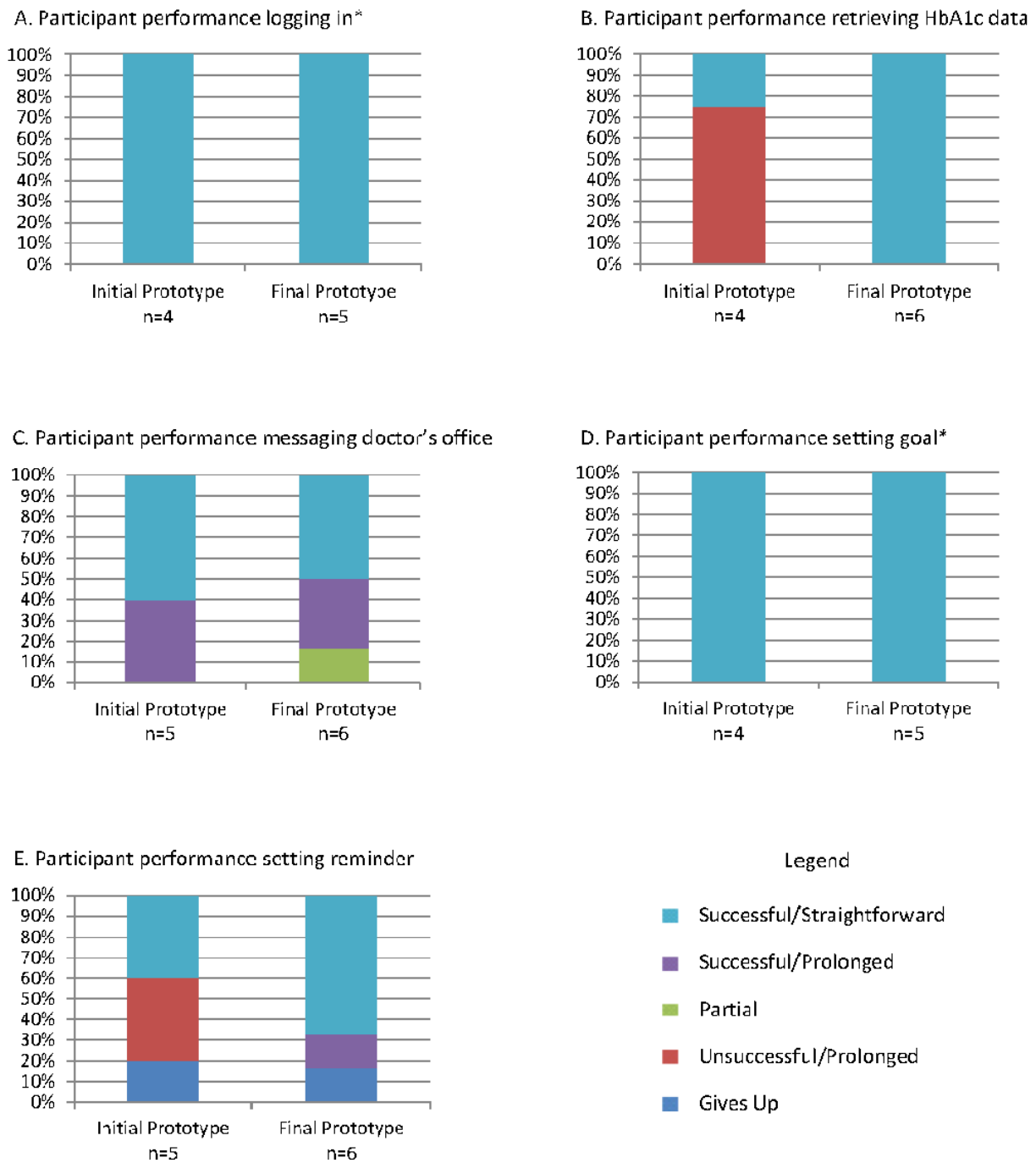
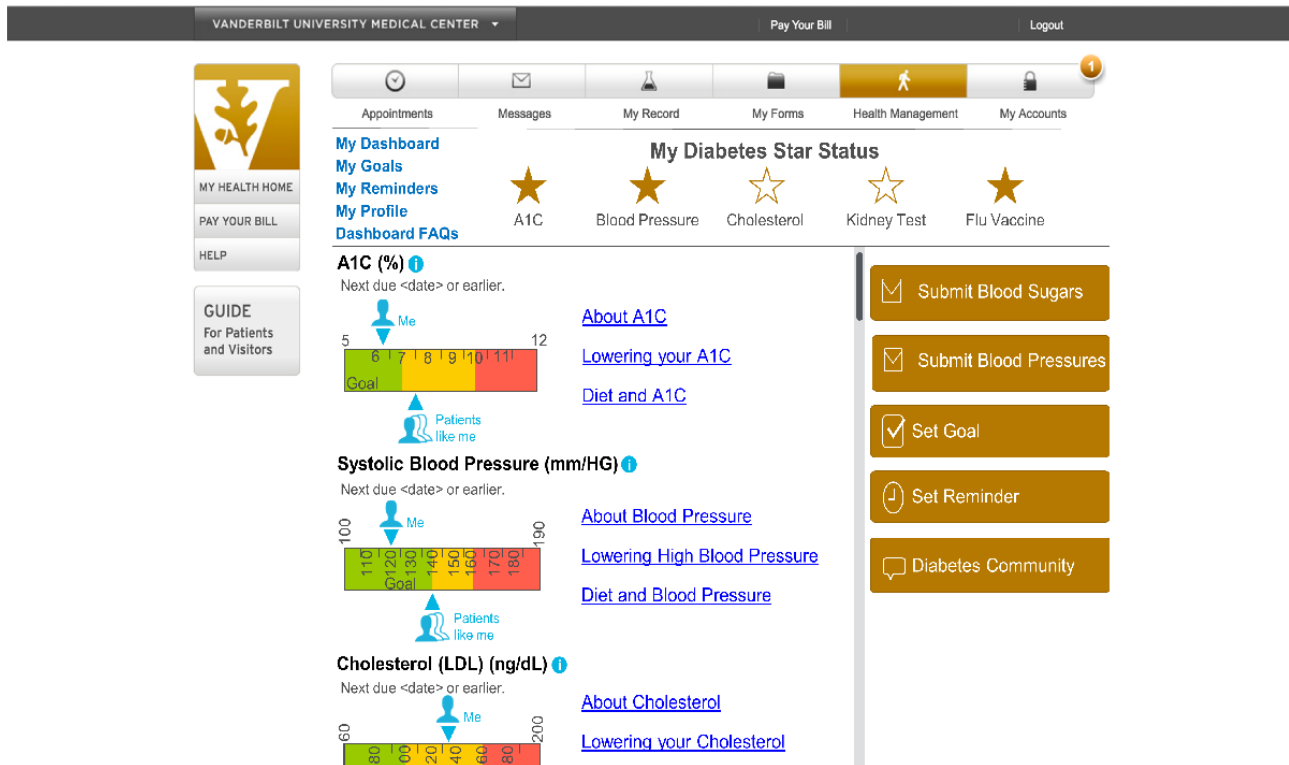


Figure 7. Screenshot of final dashboard prototype. A<sub>1c</sub>: hemoglobin A<sub>1c</sub>.



**Task: (E) Set a Reminder**

Only 2 participants in round 1 were able to set a reminder on the dashboard. Participants struggled to set the frequency of recurrence and a stop date for reminders they wished to receive only for a specified time. Subsequently, the authors revised the layout of the “set reminder” pop up window to include a clear start and stop date and time, as well as a drop-down menu to set recurrences (eg, daily, weekly, etc). After revisions, 4 of 6 participants in round 3 were able to set a reminder, with one additional participant successfully completing the task with prolonged effort.

**Participant Interviews**

Table 2 shows the participants’ comments about usability concerns grouped by usability area. Several revisions were made

in response to participants’ usability concerns, including revisions to the display of patients’ health data and star status, icons indicating the patient’s value and “patients like me” value, standardizing educational links and adding diet information, grouping and standardizing action items, enlarging the font size, and providing a frequently asked questions page (see Figure 1 initial prototype and Figure 7 final prototype).

**Satisfaction Survey**

Table 3 reports mean scores for the CSUQ items among participants in round 1 who tested the initial prototype compared with participants in round 3 who tested the final prototype. Participants who tested the initial prototype and those who tested the final prototype rated the usability above average (ie, scores >4 on a 7-point scale) for all 12 items. The mean score for all 12 items improved between the initial and final prototypes.

**Table 2.** Participants' concerns with dashboard usability.

Usability element and unique concern type	Illustrative quote
<b>Design</b>	
Font size	<i>It's very clear to me but I would definitely want to enlarge the size of the font.</i>
Patient status indicator	<i>I don't know what [the indicator] is supposed to be. Still I want to figure it out. The person's goal would be about 6.2 and the actual would be 7.5. Is that correct?</i>
Reminder functionality	<i>That's a reminder, oh! That's a clock symbol. Gotcha. It could be clearer [laughs].</i>
Patients like me indicator	<i>Not clear that this [icon] is for individual. This [icon] is for group. Up to here [group icon], just add one more figure so that will show more people.</i>
Star rating	<i>There's a star over here, on this side, but does it indicate the same thing as the star rating over here? By rating, is that telling me that I'm doing poor, good, with my goals?</i>
Hover over functionality	<i>No I wouldn't have known [I could hover over]. Once you clicked, then I realized.</i>
Goal setting functionality	<i>The end date [for the goal], you're talking about the last day of your, I don't get that. The end date [for the goal]. Help me.</i>
<b>Efficiency of use</b>	
Redundancy	<i>I mean those two things [my medical concerns drop down menu] and the message subject [free text; are the same].</i>
<b>Content and terminology</b>	
Historical values	<i>I'd actually like to see what my last three [HbA<sub>1c</sub>] were.</i>
Medical jargon	<i>I don't even know what [microalbumin] is. I've never heard of that.</i>
Diet information	<i>If you could just do something about diet. I don't see that on there anywhere. I mean, because that's like a big part of it, like what can I eat, what should I eat.</i>
Online community	<i>You're not going to be able to communicate with other patients and talk about the key things they do for support. That might be something you would add.</i>

**Table 3.** Computer system usability questionnaire survey items assessing the dashboard usability: initial versus final prototype.

Item	Initial prototype (n=5), mean (SD)	Final prototype (n=6), mean (SD)
Overall, I am satisfied with how easy it is to use this system.	5.6 (1.1)	6.3 (0.8)
It is simple to use this system.	6.0 (0.8)	6.3 (0.8)
I feel comfortable using this system.	5.7 (1.3)	6.5 (1.3)
It was easy to learn to use this system.	6.2 (0.8)	6.5 (0.8)
It is easy to find the information I need.	5.6 (1.5)	4.8 (1.2)
The information provided with the system is easy to understand.	5.4 (1.7)	5.8 (1.2)
The organization of information on the system screens is clear.	4.2 (2.2)	6.5 (0.5)
The interface of this system is pleasant.	5.4 (1.3)	6.5 (0.5)
I like using the interface of this system.	5.4 (1.1)	6.5 (0.5)
The system has all the functions and capabilities I expect it to have.	6.0 (0.7)	6.2 (0.8)
Overall, I am satisfied with this system.	5.8 (0.4)	6.7 (0.5)
The system is visually appealing.	5.8 (1.3)	6.5 (0.5)

## Discussion

### Principal Findings

Our study illustrates the use of design sprint methodology alongside mixed-methods, task-based usability testing in the design of a Web-based intervention for patients with diabetes. By using this design approach, we were able to rapidly create a prototype and rigorously assess task-based usability before

any programming. Task-based usability testing and qualitative analysis of interviews with a small number of participants quickly identified usability challenges that led to improvements in successive iterations. Participant feedback informed changes in the data display that led to improved comprehension of diabetes health data. Participants' usability satisfaction surveys demonstrated a high level of satisfaction with the dashboard that improved from initial to final prototype. The final prototype incorporated recommended strategies to enhance patient

activation across the engagement spectrum, from providing educational resources to promoting behavior change through rewards (see Figure 8) [16].

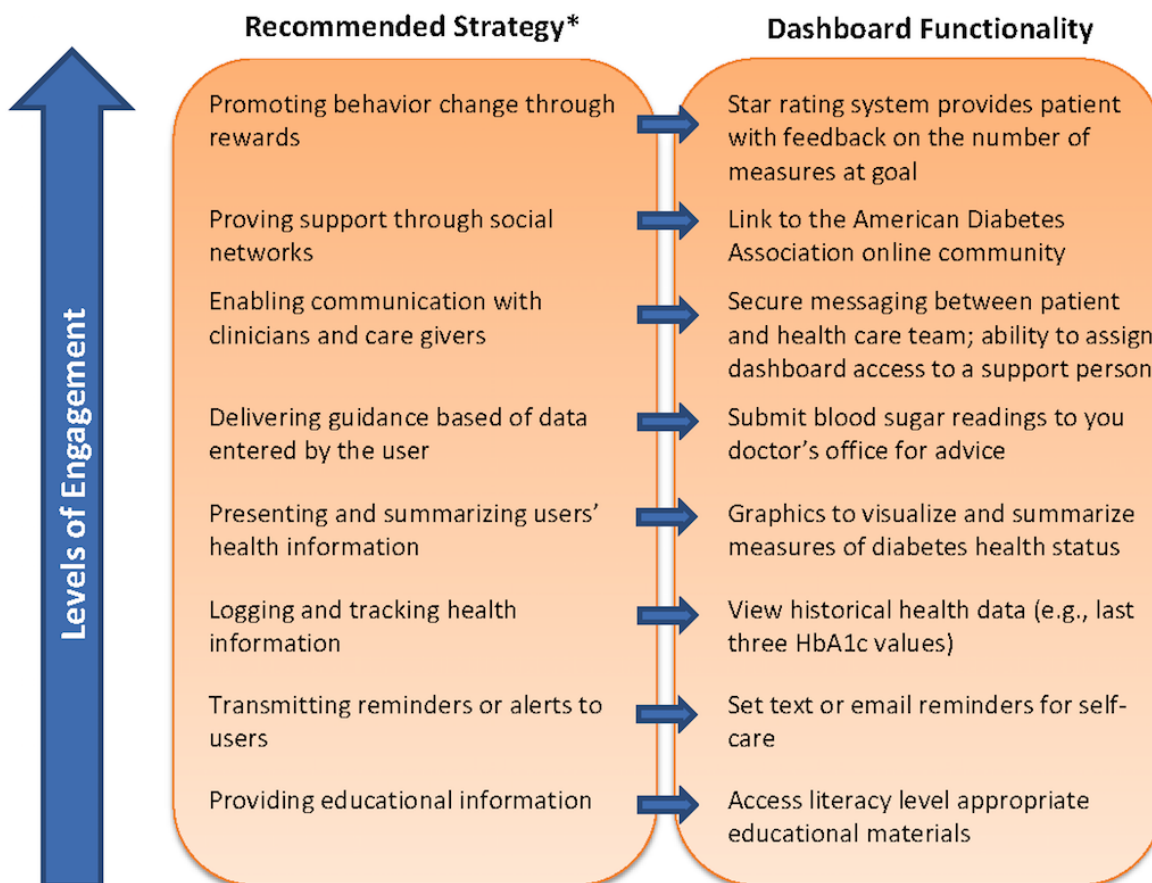
**Building Upon Prior Research**

Several prior studies have reported the design and usability of patient-facing health apps and Web-based interventions for patients with diabetes [50,53-58]. Approaches to the design of these health apps and Web-based interventions typically employ some variation of user-centered design [56-59]. A significant limitation of prior design approaches is the time and cost involved with the rapidly evolving pace of technology [60,61]. This study is the first in our knowledge to report the design of a digital health intervention using design sprint methodology and demonstrate its utility in efficiently and effectively designing a Web-based intervention that is satisfying to use.

By utilizing design sprint methodology, we were able to create a viable initial prototype within 5 days. Given the rapidly evolving technology and patient expectations of health technology [60,62], efficient yet rigorous design methodology is essential. We were able to enhance the scientific rigor of the design sprint approach by using validated measures of usability [46] and task-performance [47-49], as well as an established qualitative methodology to analyze interviews and determine saturation [51]. This approach allows usability concerns to be identified before programming, potentially saving the researcher both time and money. Consistent with the findings of Nielsen,

we found that the majority of usability problems were identified in the first 5 usability evaluations, with diminishing returns after the eighth evaluation [40-42]. While enrolling additional participants in our study may have revealed additional usability concerns, our sample was sufficient to establish a minimally viable product (eg, final prototype) that allowed us to proceed to program the dashboard with the reasonable confidence that most usability issues were identified and addressed. As with any app or website, ongoing attention to user feedback and iterative improvements are likely to continue indefinitely as technology and users evolve. Although some usability studies employ a large number of participants, this is mostly done to provide sufficient sample size for quantitative analyses, and additional participants yield relatively few new usability concerns [40-42]. In addition, our usability findings build upon other recent studies of patient-facing diabetes health apps [50,53,59]. Georgsson et al used a similar mixed-methods approach to evaluate the usability of their mHealth system for diabetes type 2 self-management [53]. Similar to this study, their study included task-based testing with a *think-aloud* protocol, semistructured interviews, and a questionnaire on patients' experiences using their system. Consistent with Georgsson et al, we found a mixed-methods approach resulted in a comprehensive understanding of usability. Our study extends these findings by demonstrating the effectiveness of this approach to objectively assess and track usability in response to iterative revisions of a prototype in the design phase.

**Figure 8.** Recommended strategies for patient activation and paired dashboard functionality by level of patient engagement. The asterisk refers to the engagement pyramid reported by Singh et al, 2016 [16]. HbA<sub>1c</sub>: hemoglobin A<sub>1c</sub>.



Our study also has implications for the design of patient portals and the display of patients' health data. By giving patients direct access to their health data, patient portals can improve patient engagement [63] and empower patients to actively participate in their care [64]. However, research suggests that patients struggle to understand health data communicated to them via patient portals [65]. A recent study by Giardian et al suggests that current patient portals do not display health data in a patient-centered way, which can lead to misunderstandings and patient distress [66]. In our study, patients had difficulty comprehending HbA<sub>1c</sub> data in the dial display (Figure 1) that improved with ruler display (Figure 7), demonstrating the importance of user-centered design. Although the content was relatively unchanged, we revised the display based on user feedback, resulting in increased comprehension and improved visibility of features aimed at facilitating patients' understanding of their health data.

### Limitations

This study has important limitations. We recruited a convenience sample of patients from a single, large, urban academic medical center that may limit the generalizability of our findings. Our sample included patients who were more educated and had greater computer and internet access than the overall population of patients with diabetes [67,68]. For future studies, researchers should consider purposive sampling to recruit patients with specific characteristics. Given the known barriers to usability among older patients [15], a strength of our sample was the inclusion of a majority of patients over the age of 60 years that allowed us to ensure the dashboard usability among this demographic. In addition, although we were able to directly observe individual users as they attempted several assigned tasks using the dashboard, our data are subject to the Hawthorne effect (ie, altered behavior due to an awareness of being observed). Similarly, we did not collect data on how patients would engage with the dashboard on their own. It would be

useful to collect *actual-use* data in future studies including the level of engagement with specific dashboard functions over time. Although we designed the dashboard with elements aimed at increasing patient activation, this study focused on the design and task-based usability of the dashboard and not on the evaluation of its impact. Further research is needed to test the efficacy of the dashboard on cognitive, behavioral, and clinical outcomes including patient activation.

Researchers and others considering using design sprint methodology should also consider some of the limitations of the approach. Although a standard design sprint that unfolds over 5 days is generally recommended [17,18], researchers may wish to experiment with shorter, or more likely, longer sprints. Design sprint methodology relies on understanding the user (ie, the consumer and their needs), and in some instances, it may be necessary to spend additional time before the design sprint to understand the target user and their needs and challenges. In our case, a literature review on the patients' experiences with portal use, challenges with diabetes self-management, and the limitations of existing diabetes apps provided insights about our target users. Design sprints also rely heavily on the ideas generated from the solutions sketched by team members on day 2. Therefore, this phase of idea generation should not be shortened and may, in fact, benefit from more time.

### Conclusions

In conclusion, the results underscore the value of design sprint methodology to efficiently create a viable user-centric prototype of a Web-based intervention and the importance of mixed-methods evaluation of usability as a part of the design phase beginning with the initial prototype. Design sprints offer an efficient way to define the problem, assess the needs of users, iteratively generate ideas and develop a viable product for testing, whereas usability evaluation methods ensure health apps and Web-based interventions appeal to users and support their use.

---

### Acknowledgments

This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases / National Institutes of Health (K23DK106511 and 2P30DK092986-07) and the National Center for Advancing Translational Sciences / National Institutes of Health UL1 TR000445. We are grateful to Ricardo J Trochez, BA, and Kemberlee R Bonnet, MA, for their assistance as coders.

---

### Conflicts of Interest

None declared.

---

### References

1. Centers for Disease Control and Prevention. Atlanta, GA: US Department of Health and Human Services; 2017. National Diabetes Statistics Report, 2017 URL: <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf> [WebCite Cache ID 713Xcg271]
2. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes self-management education and support in type 2 diabetes. *Diabetes Educ* 2017 Dec;43(1):40-53. [doi: [10.1177/0145721716689694](https://doi.org/10.1177/0145721716689694)] [Medline: [28118121](https://pubmed.ncbi.nlm.nih.gov/28118121/)]
3. Shrivastava SR, Shrivastava PS, Ramasamy J. Role of self-care in management of diabetes mellitus. *J Diabetes Metab Disord* 2013 Mar 05;12(1):14 [FREE Full text] [doi: [10.1186/2251-6581-12-14](https://doi.org/10.1186/2251-6581-12-14)] [Medline: [23497559](https://pubmed.ncbi.nlm.nih.gov/23497559/)]
4. Bennett KJ, McDermott S, Mann JR, Hardin J. Receipt of recommended services among patients with selected disabling conditions and diabetes. *Disabil Health J* 2017 Jan;10(1):58-64. [doi: [10.1016/j.dhjo.2016.09.001](https://doi.org/10.1016/j.dhjo.2016.09.001)] [Medline: [27746151](https://pubmed.ncbi.nlm.nih.gov/27746151/)]



5. McBrien KA, Naugler C, Ivers N, Weaver RG, Campbell D, Desveaux L, et al. Barriers to care in patients with diabetes and poor glycemic control: a cross-sectional survey. *PLoS One* 2017 May 1;12(5):e0176135 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0176135](https://doi.org/10.1371/journal.pone.0176135)] [Medline: [28459820](#)]
6. Sacks RM, Greene J, Hibbard J, Overton V, Parrotta CD. Does patient activation predict the course of type 2 diabetes? A longitudinal study. *Patient Educ Couns* 2017 Jul;100(7):1268-1275. [doi: [10.1016/j.pec.2017.01.014](https://doi.org/10.1016/j.pec.2017.01.014)] [Medline: [28159442](#)]
7. Bolen SD, Chandar A, Falck-Ytter C, Tyler C, Perzynski AT, Gertz AM, et al. Effectiveness and safety of patient activation interventions for adults with type 2 diabetes: systematic review, meta-analysis, and meta-regression. *J Gen Intern Med* 2014 Aug;29(8):1166-1176 [[FREE Full text](#)] [doi: [10.1007/s11606-014-2855-4](https://doi.org/10.1007/s11606-014-2855-4)] [Medline: [24733301](#)]
8. Sullivan SD, Dalal MR, Burke JP. The impact of diabetes counseling and education: clinical and cost outcomes from a large population of US managed care patients with type 2 diabetes. *Diabetes Educ* 2013;39(4):523-531. [doi: [10.1177/0145721713486525](https://doi.org/10.1177/0145721713486525)] [Medline: [23640302](#)]
9. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR Mhealth Uhealth* 2017 Mar 14;5(3):e35 [[FREE Full text](#)] [doi: [10.2196/mhealth.6522](https://doi.org/10.2196/mhealth.6522)] [Medline: [28292740](#)]
10. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. *J Diabetes Complications* 2014;28(2):243-251 [[FREE Full text](#)] [doi: [10.1016/j.jdiacomp.2013.07.003](https://doi.org/10.1016/j.jdiacomp.2013.07.003)] [Medline: [24332469](#)]
11. Yu CH, Bahniwal R, Laupacis A, Leung E, Orr MS, Straus SE. Systematic review and evaluation of web-accessible tools for management of diabetes and related cardiovascular risk factors by patients and healthcare providers. *J Am Med Inform Assoc* 2012;19(4):514-522 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000307](https://doi.org/10.1136/amiajnl-2011-000307)] [Medline: [22215057](#)]
12. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. *J Med Internet Res* 2011 Jan 25;13(1):e9 [[FREE Full text](#)] [doi: [10.2196/jmir.1391](https://doi.org/10.2196/jmir.1391)] [Medline: [21371992](#)]
13. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. *Obesity (Silver Spring)* 2010 Feb;18(2):308-313 [[FREE Full text](#)] [doi: [10.1038/oby.2009.242](https://doi.org/10.1038/oby.2009.242)] [Medline: [19696764](#)]
14. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. *Biomed Instrum Technol* 2012;Suppl:49-56. [doi: [10.2345/0899-8205-46.s2.49](https://doi.org/10.2345/0899-8205-46.s2.49)] [Medline: [23039777](#)]
15. Isaković M, Sedlar U, Volk M, Bešter J. Usability pitfalls of diabetes mHealth apps for the elderly. *J Diabetes Res* 2016;2016:1604609 [[FREE Full text](#)] [doi: [10.1155/2016/1604609](https://doi.org/10.1155/2016/1604609)] [Medline: [27034957](#)]
16. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. *Issue Brief (Commonw Fund)* 2016 Feb;5:1-11. [Medline: [26934758](#)]
17. Banfield R, Lombardo CT, Wax T. *Design Sprint: A Practical Guidebook for Building Great Digital Products*. Sebastopol, CA: O'Reilly Media, Inc; 2015.
18. Knapp J, Zeratsky J, Kowitz B. *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days*. New York, NY: Simon and Schuster; 2016.
19. Osborn CY, Rosenbloom ST, Stenner SP, Anders S, Muse S, Johnson KB, et al. MyHealthAtVanderbilt: policies and procedures governing patient portal functionality. *J Am Med Inform Assoc* 2011 Dec;18 Suppl 1:i18-i23 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000184](https://doi.org/10.1136/amiajnl-2011-000184)] [Medline: [21807648](#)]
20. Sweileh WM, Zyoud SH, Abu Nab'a RJ, Deleq MI, Enaia MI, Nassar SM, et al. Influence of patients' disease knowledge and beliefs about medicines on medication adherence: findings from a cross-sectional survey among patients with type 2 diabetes mellitus in Palestine. *BMC Public Health* 2014 Jan 30;14:94 [[FREE Full text](#)] [doi: [10.1186/1471-2458-14-94](https://doi.org/10.1186/1471-2458-14-94)] [Medline: [24479638](#)]
21. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. *Diabetes Technol Ther* 2014 Nov;16(11):784-793. [doi: [10.1089/dia.2014.0078](https://doi.org/10.1089/dia.2014.0078)] [Medline: [24999599](#)]
22. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. *Pediatr Diabetes* 2013 Jun;14(4):231-238 [[FREE Full text](#)] [doi: [10.1111/pedi.12034](https://doi.org/10.1111/pedi.12034)] [Medline: [23627878](#)]
23. Kuo A, Dang S. Secure messaging in electronic health records and its impact on diabetes clinical outcomes: a systematic review. *Telemed J E Health* 2016 Dec;22(9):769-777. [doi: [10.1089/tmj.2015.0207](https://doi.org/10.1089/tmj.2015.0207)] [Medline: [27027337](#)]
24. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. *Curr Diab Rep* 2010 Dec;10(6):422-435 [[FREE Full text](#)] [doi: [10.1007/s11892-010-0151-1](https://doi.org/10.1007/s11892-010-0151-1)] [Medline: [20890688](#)]
25. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. *JMIR Mhealth Uhealth* 2015 Feb 26;3(1):e20 [[FREE Full text](#)] [doi: [10.2196/mhealth.3335](https://doi.org/10.2196/mhealth.3335)] [Medline: [25803705](#)]
26. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. *J Med Internet Res* 2013 Jun 14;15(6):e120 [[FREE Full text](#)] [doi: [10.2196/jmir.2600](https://doi.org/10.2196/jmir.2600)] [Medline: [23770578](#)]

27. Schokker MC, Keers JC, Bouma J, Links TP, Sanderman R, Wolffenbuttel BH, et al. The impact of social comparison information on motivation in patients with diabetes as a function of regulatory focus and self-efficacy. *Health Psychol* 2010 Jul;29(4):438-445. [doi: [10.1037/a0019878](https://doi.org/10.1037/a0019878)] [Medline: [20658832](https://pubmed.ncbi.nlm.nih.gov/20658832/)]
28. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. *J Diabetes Sci Technol* 2013 Jan 01;7(1):247-262 [FREE Full text] [doi: [10.1177/193229681300700130](https://doi.org/10.1177/193229681300700130)] [Medline: [23439183](https://pubmed.ncbi.nlm.nih.gov/23439183/)]
29. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. *J Behav Med* 2016 Dec;39(6):981-994. [doi: [10.1007/s10865-016-9765-3](https://doi.org/10.1007/s10865-016-9765-3)] [Medline: [27412774](https://pubmed.ncbi.nlm.nih.gov/27412774/)]
30. Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. *Curr Diab Rep* 2014 Oct;14(10):534 [FREE Full text] [doi: [10.1007/s11892-014-0534-9](https://doi.org/10.1007/s11892-014-0534-9)] [Medline: [25173689](https://pubmed.ncbi.nlm.nih.gov/25173689/)]
31. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. *J Med Internet Res* 2014 Apr 09;16(4):e104 [FREE Full text] [doi: [10.2196/jmir.2968](https://doi.org/10.2196/jmir.2968)] [Medline: [24718852](https://pubmed.ncbi.nlm.nih.gov/24718852/)]
32. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). *J Health Commun* 2010;15 Suppl 2:183-196 [FREE Full text] [doi: [10.1080/10810730.2010.499988](https://doi.org/10.1080/10810730.2010.499988)] [Medline: [20845203](https://pubmed.ncbi.nlm.nih.gov/20845203/)]
33. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. *Health Serv Res* 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: [10.1111/j.1475-6773.2004.00269.x](https://doi.org/10.1111/j.1475-6773.2004.00269.x)] [Medline: [15230939](https://pubmed.ncbi.nlm.nih.gov/15230939/)]
34. Martinez W, Wallston KA, Schlundt DG, Hickson GB, Bonnet KR, Trochez RJ, et al. Patients' perspectives on social and goal-based comparisons regarding their diabetes health status. *Br Med J Open Diabetes Res Care* 2018;6(1):e000488 [FREE Full text] [doi: [10.1136/bmjdr-2017-000488](https://doi.org/10.1136/bmjdr-2017-000488)] [Medline: [29862032](https://pubmed.ncbi.nlm.nih.gov/29862032/)]
35. Siddiqui H. Keynote Animation – How To Prototype UI. Freiburg, Germany: Smashing Magazine; 2015. URL: <https://www.smashingmagazine.com/2015/08/animating-in-keynote/> [WebCite Cache ID 70v0mlyCA]
36. Elder NC, Barney K. “But what does it mean for me?” Primary care patients' communication preferences for test results notification. *Jt Comm J Qual Patient Saf* 2012 Apr;38(4):168-176. [Medline: [22533129](https://pubmed.ncbi.nlm.nih.gov/22533129/)]
37. Torsvik T, Lillebo B, Mikkelsen G. Presentation of clinical laboratory results: an experimental comparison of four visualization techniques. *J Am Med Inform Assoc* 2013;20(2):325-331 [FREE Full text] [doi: [10.1136/amiainl-2012-001147](https://doi.org/10.1136/amiainl-2012-001147)] [Medline: [23043123](https://pubmed.ncbi.nlm.nih.gov/23043123/)]
38. Zikmund-Fisher BJ, Scherer AM, Witteman HO, Solomon JB, Exe NL, Tarini BA, et al. Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *J Am Med Inform Assoc* 2017 May 1;24(3):520-528. [Medline: [28040686](https://pubmed.ncbi.nlm.nih.gov/28040686/)]
39. National Institutes of Health. 2007. Guidelines for the Diagnosis and Management of Asthma URL: <https://www.nhlbi.nih.gov/files/docs/guidelines/asthsumm.pdf> [WebCite Cache ID 713WLNKja]
40. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. 1993 Presented at: INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213.
41. Nielsen J. Usability inspection methods. 1994 Presented at: CHI 94 ACM Conference Companion on Human Factors in Computing Systems; April 24-28, 1994; Boston, MA p. 413-414. [doi: [10.1145/259963.260531](https://doi.org/10.1145/259963.260531)]
42. Nielsen J. Nielsen Norman Group. 2000. Why You Only Need to Test with 5 Users URL: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [WebCite Cache ID 6x8Whgxdh]
43. Sarkar U, Schillinger D, López A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. *J Gen Intern Med* 2011 Mar;26(3):265-271 [FREE Full text] [doi: [10.1007/s11606-010-1552-1](https://doi.org/10.1007/s11606-010-1552-1)] [Medline: [21057882](https://pubmed.ncbi.nlm.nih.gov/21057882/)]
44. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Making* 2007;27(5):672-680. [Medline: [17641137](https://pubmed.ncbi.nlm.nih.gov/17641137/)]
45. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. *Int J Med Inform* 2004 Nov;73(11-12):781-795. [doi: [10.1016/j.ijmedinf.2004.08.003](https://doi.org/10.1016/j.ijmedinf.2004.08.003)] [Medline: [15491929](https://pubmed.ncbi.nlm.nih.gov/15491929/)]
46. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int J Hum-Comput Interact* 1995 Jan;7(1):57-78. [doi: [10.1080/10447319509526110](https://doi.org/10.1080/10447319509526110)]
47. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. *J Gen Intern Med* 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: [10.1007/s11606-016-3771-6](https://doi.org/10.1007/s11606-016-3771-6)] [Medline: [27418347](https://pubmed.ncbi.nlm.nih.gov/27418347/)]
48. Taha J, Sharit J, Czaja SJ. The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. *J Appl Gerontol* 2014 Jun;33(4):416-436 [FREE Full text] [doi: [10.1177/0733464812447283](https://doi.org/10.1177/0733464812447283)] [Medline: [24781964](https://pubmed.ncbi.nlm.nih.gov/24781964/)]
49. Segall N, Saville JG, L'Engle P, Carlson B, Wright MC, Schulman K, et al. Usability evaluation of a personal health record. *AMIA Annu Symp Proc* 2011;2011:1233-1242 [FREE Full text] [Medline: [22195184](https://pubmed.ncbi.nlm.nih.gov/22195184/)]

50. Nelson LA, Bethune MC, Lagotte AE, Osborn CY. The usability of diabetes MAP: a web-delivered intervention for improving medication adherence. *JMIR Hum Factors* 2016 May 12;3(1):e13 [FREE Full text] [doi: [10.2196/humanfactors.5177](https://doi.org/10.2196/humanfactors.5177)] [Medline: [27174496](https://pubmed.ncbi.nlm.nih.gov/27174496/)]
51. Strauss A, Corbin J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Newbury Park, CA: Sage Publications; 1990.
52. US Department of Health Human Services. Usability.gov. 2014. Usability evaluation basics URL: <https://www.usability.gov/what-and-why/usability-evaluation.html> [WebCite Cache ID 713VoTuTO]
53. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. *J Biomed Inform* 2016 Feb;59:115-129 [FREE Full text] [doi: [10.1016/j.jbi.2015.11.008](https://doi.org/10.1016/j.jbi.2015.11.008)] [Medline: [26639894](https://pubmed.ncbi.nlm.nih.gov/26639894/)]
54. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *J Am Med Inform Assoc* 2016;23(1):5-11 [FREE Full text] [doi: [10.1093/jamia/ocv099](https://doi.org/10.1093/jamia/ocv099)] [Medline: [26377990](https://pubmed.ncbi.nlm.nih.gov/26377990/)]
55. Fu H, McMahan SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. *Diabetes Res Clin Pract* 2017 Sep;131:70-81. [doi: [10.1016/j.diabres.2017.06.016](https://doi.org/10.1016/j.diabres.2017.06.016)] [Medline: [28692830](https://pubmed.ncbi.nlm.nih.gov/28692830/)]
56. Grant RW, Wald JS, Poon EG, Schnipper JL, Gandhi TK, Volk LA, et al. Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: patient gateway for diabetes collaborative care. *Diabetes Technol Ther* 2006 Oct;8(5):576-586 [FREE Full text] [doi: [10.1089/dia.2006.8.576](https://doi.org/10.1089/dia.2006.8.576)] [Medline: [17037972](https://pubmed.ncbi.nlm.nih.gov/17037972/)]
57. Osborn CY, Mulvaney SA. Development and feasibility of a text messaging and interactive voice response intervention for low-income, diverse adults with type 2 diabetes mellitus. *J Diabetes Sci Technol* 2013 May 01;7(3):612-622 [FREE Full text] [doi: [10.1177/193229681300700305](https://doi.org/10.1177/193229681300700305)] [Medline: [23759393](https://pubmed.ncbi.nlm.nih.gov/23759393/)]
58. Yu CH, Parsons JA, Hall S, Newton D, Jovicic A, Lottridge D, et al. User-centered design of a web-based self-management site for individuals with type 2 diabetes-providing a sense of control and community. *BMC Med Inform Decis Mak* 2014 Jul 23;14:60 [FREE Full text] [doi: [10.1186/1472-6947-14-60](https://doi.org/10.1186/1472-6947-14-60)] [Medline: [25056379](https://pubmed.ncbi.nlm.nih.gov/25056379/)]
59. Alanzi T, Istepanian R, Philip N. Design and usability evaluation of social mobile diabetes management system in the gulf region. *JMIR Res Protoc* 2016 Sep 26;5(3):e93 [FREE Full text] [doi: [10.2196/resprot.4348](https://doi.org/10.2196/resprot.4348)] [Medline: [27670696](https://pubmed.ncbi.nlm.nih.gov/27670696/)]
60. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. *Am J Prev Med* 2016 Nov;51(5):816-824. [doi: [10.1016/j.amepre.2016.05.001](https://doi.org/10.1016/j.amepre.2016.05.001)] [Medline: [27745681](https://pubmed.ncbi.nlm.nih.gov/27745681/)]
61. Da Silva TS, Martin A, Maurer F, Silveira M. User-centered design and agile methods: a systematic review. 2011 Presented at: Agile Conference (AGILE); August 7-13, 2011; Salt Lake City, UT p. 77-86. [doi: [10.1109/AGILE.2011.24](https://doi.org/10.1109/AGILE.2011.24)]
62. Lithgow K, Edwards A, Rabi D. Smartphone app use for diabetes management: evaluating patient perspectives. *JMIR Diabetes* 2017 Jan 23;2(1):e2. [doi: [10.2196/diabetes.6643](https://doi.org/10.2196/diabetes.6643)]
63. Tulu B, Trudel J, Strong DM, Johnson SA, Sundaresan D, Garber L. Patient portals: an underused resource for improving patient engagement. *Chest* 2016 Jan;149(1):272-277. [doi: [10.1378/chest.14-2559](https://doi.org/10.1378/chest.14-2559)] [Medline: [26066707](https://pubmed.ncbi.nlm.nih.gov/26066707/)]
64. Archer N, Fevrier-Thomas U, Lokker C, McKibbin KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc* 2011;18(4):515-522 [FREE Full text] [doi: [10.1136/amiainl-2011-000105](https://doi.org/10.1136/amiainl-2011-000105)] [Medline: [21672914](https://pubmed.ncbi.nlm.nih.gov/21672914/)]
65. Giardina TD, Modi V, Parrish DE, Singh H. The patient portal and abnormal test results: an exploratory study of patient experiences. *Patient Exp J* 2015;2(1):148-154 [FREE Full text] [Medline: [28345018](https://pubmed.ncbi.nlm.nih.gov/28345018/)]
66. Giardina TD, Baldwin J, Nystrom DT, Sittig DF, Singh H. Patient perceptions of receiving test results via online portals: a mixed-methods study. *J Am Med Inform Assoc* 2018 Apr 01;25(4):440-446 [FREE Full text] [doi: [10.1093/jamia/ocx140](https://doi.org/10.1093/jamia/ocx140)] [Medline: [29240899](https://pubmed.ncbi.nlm.nih.gov/29240899/)]
67. Geiss LS, Wang J, Cheng YJ, Thompson TJ, Barker L, Li Y, et al. Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980-2012. *J Am Med Assoc* 2014 Sep 24;312(12):1218-1226. [doi: [10.1001/jama.2014.11494](https://doi.org/10.1001/jama.2014.11494)] [Medline: [25247518](https://pubmed.ncbi.nlm.nih.gov/25247518/)]
68. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. *Med Care* 2012 May;50(5):434-440. [doi: [10.1097/MLR.0b013e318249d81b](https://doi.org/10.1097/MLR.0b013e318249d81b)] [Medline: [22354209](https://pubmed.ncbi.nlm.nih.gov/22354209/)]

## Abbreviations

- CSUQ:** Computer System Usability Questionnaire
- HbA<sub>1c</sub>:** hemoglobin A1c
- HMW:** how might we
- MHAV:** My Health at Vanderbilt
- UX:** user experience
- VAPC:** Vanderbilt Adult Primary Care
- VUMC:** Vanderbilt University Medical Center

*Edited by G Eysenbach; submitted 04.12.17; peer-reviewed by GL Dimaguila, J Waycott; comments to author 17.03.18; revised version received 17.07.18; accepted 17.07.18; published 24.09.18.*

*Please cite as:*

*Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA*

*A Patient-Facing Diabetes Dashboard Embedded in a Patient Web Portal: Design Sprint and Usability Testing*

*JMIR Hum Factors 2018;5(3):e26*

*URL: <http://humanfactors.jmir.org/2018/3/e26/>*

*doi: [10.2196/humanfactors.9569](https://doi.org/10.2196/humanfactors.9569)*

*PMID: [30249579](https://pubmed.ncbi.nlm.nih.gov/30249579/)*

©William Martinez, Anthony L Threatt, S Trent Rosenbloom, Kenneth A Wallston, Gerald B Hickson, Tom A Elasy. Originally published in JMIR Human Factors (<http://humanfactors.jmir.org>), 24.09.2018. 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 <http://humanfactors.jmir.org>, as well as this copyright and license information must be included.

---

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

---

<https://www.jmirpublications.com/>