JMIR Human Factors

Making health care interventions and technologies usable, safe, and effective
Volume 9 (2022), Issue 4 ISSN: 2292-9495

Contents

Original Papers

Reviews

Original Paper

Knowledge, Attitudes, and Perceptions of Jordanians Toward Adopting and Using Telemedicine: National Cross-sectional Study

Rand Murshidi[1*], MD; Muhammad Hammouri[2*], MD; Hana Taha[3], MD, PhD; Razi Kitaneh[2], MD; Mahmoud Alshneikat[2], MD; Abdallah Al-Qawasmeh[2], MD; Ahmad Al-Oleimat[2], MD; Leen Al-Huneidy[2,4], MD; Yazan Al-Huneidy[2], MD; Abdallah Al-Ani[5], MD

[1]Department of Dermatology, School of Medicine, The University of Jordan, Amman, Jordan

[2]School of Medicine, The University of Jordan, Amman, Jordan

[3]Department of Basic Medical Sciences, The Hashemite University, Zarqa, Jordan

[4]Department of Internal Medicine, King Hussein Medical Center, Amman, Jordan

[5]Office of Scientific Affairs and Research, King Hussein Cancer Center, Amman, Jordan

[*]these authors contributed equally

Corresponding Author:
Abdallah Al-Ani, MD
Office of Scientific Affairs and Research
King Hussein Cancer Center
Queen Rania Street
Amman, 11194
Jordan
Phone: 962 791184320
Email: abdallahalany@gmail.com

Abstract

Background: Due to the upsurge of COVID-19, nations are increasingly adopting telemedicine programs in anticipation of similar crises. Similar to all nations worldwide, Jordan is implementing efforts to adopt such technologies, yet it is far from complete.

Objective: This study aims to assess the knowledge, attitudes, and perceptions of Jordanians toward telemedicine, to identify key factors predisposing individuals to its use or acting as barriers to its implementation.

Methods: We implemented a cross-sectional design using an online, self-administered questionnaire executed in Google Forms and distributed through social media. Differences in knowledge and attitude scores were examined using independent sample *t* tests and ANOVA. A multivariate linear regression model was computed to assess predictors of awareness toward telemedicine.

Results: A total of 1201 participants fully completed the questionnaire. Participants were characterized by a mean age of 36.3 (SD 14.4) years and a male-to-female ratio of nearly 1:1. About 50% (619/1201, 51.5%) of our studied population were aware of telemedicine, while nearly 25% (299/1201, 24.9%) declared they had observed it in action. Approximatively 68% (814/1201, 67.8%) of respondents were willing to use telemedicine. The majority of the sample portrayed favorable and positive views toward telemedicine. Higher educational degrees, living in urban districts, and having a higher perception of electronic usage ability were associated with higher knowledge and better attitudes toward telemedicine (all $P<.05$). The multivariate linear regression analysis demonstrated that perceived ability to use electronics was associated with positive attitudes (β=0.394; 95% CI 0.224 to 0.563), while living in Southern Jordan predicted poor attitudes toward telemedicine (β=–2.896; 95% CI –4.873 to –0.919).

Conclusions: Jordanians portray favorable perceptions of telemedicine. Nonetheless, concerns with regards to privacy, medical errors, and capacity for accurate diagnoses are prevalent. Furthermore, Jordanians believe that integrating telemedicine within the health care system is not applicable due to limited resources.

(JMIR Hum Factors 2022;9(4):e41499) doi:10.2196/41499

KEYWORDS

Jordan; telemedicine; KAP study; COVID-19; telehealth; regression model; user perception; online survey; privacy; health care system

Introduction

Telemedicine is defined by the World Health Organization as the use of information and communication technologies to promote health, provide medical care, exchange medical information, and educate health care providers and patients over long distances [1]. It was properly integrated into clinical practice as early as 1959 within the field of psychiatry [2]. In recent years, there has been expeditious adoption and rising interest by policy makers of this technology across most disciplines of medical care worldwide [3-5]. Those changes were a direct cause of and paralleled by the COVID-19 pandemic, as the need for ingenious ways to effectively provide health care while maintaining implementations of social distancing and policies to limit the spread of the virus was imperative [6].

Although the literature recognizes the myriad of benefits of telemedicine, such as reducing travel time, decreasing consultation fees, and increasing access to medical services to residents of remote, low-resource regions [7-9], impediments to the proper implementation of this technology worldwide still exist, especially within low-resource settings. Namely, hurdles related to cost, lack of technical staff, privacy of information, and resistance to change were among the most commonly reported [10-12].

Along those challenges, there are some aspects that are particularly important in the Arab context, specifically those pertaining to social and religious restrictions [13,14]. One example of this is the case of dealing with the opposite sex (ie, when a male health care provider has to cooperate with a female clinical staff or patient and vice versa). Although permitted in Islam under specific rules, the traditional beliefs of parties involved may prohibit the use of telemedicine services. Such a phenomenon is considered one of the top challenges of implanting telemedicine in Saudi health care [15].

In the Arab world, the concept of telemedicine is still relatively new, and applications of this technology are still limited [16,17]. Jordan as a country is no different to this situation as early attempts to carry out telemedicine projects of limited reach date back to 2008 [18]. Other small-scale projects, including digital mental health services and counseling of pediatric asthma patients, also took place before and in response to the COVID-19 pandemic [19-22].

To date, some of the well-established telemedicine programs in Jordan include the Hakeem Portal, which provides patients with the service of booking outpatient follow-up consultations within public health institutions [23]. Other programs include the King Hussein Cancer Center programs for retinoblastoma diagnosis and management as well as consultation services for general cancer cases [24,25], specialized telehealth consultations in some of the private sector hospitals [26], and finally, independent service providers like that of “AlTibbi,” a mobile health service [27]. Nonetheless, successful nationwide adoption of telemedicine is still far from complete. Taking into consideration the reciprocal nature of this service that requires good technological literacy, supporting infrastructure, an adequate level of patient education, and resistance to change in both staff and patients [28], successful implementation requires a thorough understanding of its tenants and evaluation of the skills required from parties involved, specifically, the recipients and providers, and an assessment of the possibility of incorporating it into the existing system, all of which has not been addressed in the Jordanian population.

Hence, the primary aim of this study was to investigate the knowledge, attitudes, and perceptions of the Jordanian public toward telemedicine. The secondary aims of this study were to identify key predictive factors of telemedicine usage and the potential obstacles for implementation.

Methods

Study Setting, Design, and Sampling

This cross-sectional study was conducted in The Hashemite Kingdom of Jordan, an upper-middle-income country located in the Middle East. Jordan has a population of 10.3 million and a median age of 23.8 years as of 2022. We designed and distributed an online, self-administered questionnaire executed in Google Forms and distributed through social media. We adopted a convenience sampling technique in order to approach the target group of the study, which aimed to represent the adult general population (≥18 years old) across all genders, governorates, and nationalities of those living in Jordan. The questionnaire was distributed among the target population during January 2022 through multiple popular social media platforms (eg, Facebook, Twitter, and WhatsApp). Participants were encouraged to share the questionnaire among their friends and relatives for maximum reach and generalizability of results. Simultaneously, a printed self-administered version of the questionnaire was utilized and distributed in places where different groups of the target population, especially those who are less likely to be active on the aforementioned social media platforms, could be found, thus reducing biases introduced by online surveys. These locations included the Jordan University Hospital, King Abdullah University Hospital, Al-Bashir public hospital, and the Royal Medical Services, as these locations provide medical care to the greater majority of individuals in Jordan. Moreover, in-person data collection was initiated in public spaces including malls, shopping districts, and entertainment districts. It should be noted that the Jordan University Hospital is the largest academic center in Jordan and is the largest referral center for all of central Jordan serving over 4 million patients. Participants included in the study were those who gave informed consent, completed the whole questionnaire, and were ≥18 years of age.

Questionnaire Development

The questionnaire was constructed after conducting an extensive literature review [29-33]. The questionnaire's first domain inquired about relevant sociodemographic variables of potential participants. These factors included age, biological sex, educational level, governorate or residence, place of residence (rural/urban), marital status, number of children, type of insurance, job type (office, field, student, and retired), type of internet connection at home, presence of chronic illness, presence of regularly administered medications, and the type of hospital the patient attends for regular follow up (governmental, private, university hospital, military, or others, if applicable). Also, the first domain examined the perceived level of competency with handling electronics using a 10-point scale. The second and third domains were primarily influenced by an exploration of attitudes toward telemedicine in Saudi Arabia [30]. The second domain was concerned with assessing the participants' level of knowledge regarding telemedicine and its services. It consisted of 8 dichotomous items with a score ranging from 0 to 8. Using the Kuder-Richardson formula, the alpha value for the knowledge domain was 0.677. The third and final domain consisted of 14 items that were used to assess the attitudes (ie, general perceptions, perceived benefits, and concerns) of the study participants toward the use of telemedicine and its applications. Each item was associated with a 4-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). A neutral option was removed as means to avoid neutral bias when using the Likert scale and ease the process of statistical analysis. The first 11 items of the third domain contributed to the attitudes score, ranging from 11 to 44. Similarly, the final 3 items of the third domain constituted the perceptions score, ranging from 3 to 12.

During pilot testing, the Cronbach alpha for the attitudes and perceptions scores were 0.82 and 0.74, respectively, thus ensuring proper internal consistency. The questionnaire's content validity was ensured by a panel of experts in global and public health. Meanwhile, face validity was ensured through respondents' feedback during pilot testing. Construct validity was examined using factor analysis. The Barret test for sphericity was significant at $P<.001$, and the Kaiser Meyer Olkin measure was 0.886. Using principal component analysis with direct oblimin rotation, a total of 6 factors had eigenvalues greater than 1 and explained 58.2% of the variance. Scree plot inspection demonstrated the point of inflexion after 3 components explained 42.7% of the variance. The other components were excluded as they managed to explain less than 5% of the variance. The 3 components corresponded to the knowledge, attitude, and perception domains.

The questionnaire was translated to Arabic to ensure comprehensibility, and it was translated back to English, all through the help of an expert translator. The final questionnaire included a total of 36 items (including demographics), with an approximate time of completion of 4 minutes.

Statistical Analysis

Statistical Methods

Statistical analysis was conducted using SPSS version 24 (IBM Corp, Armonk, NY). Data were reported as frequencies (n) and percentages (%) or means (SDs) wherever applicable. Normality of data was tested using the Shapiro-Wilk and Kolmogorov Smirnov tests. Mean differences were examined using the independent sample *t* test or ANOVA. A multivariate linear regression model was computed to assess the predictors of attitudes score. All statistical tests were conducted with 95% CIs and a 5% error margin. A P value <.05 was considered statistically significant.

Power and Sample Size

Considering that there are no comparable studies from Jordan that can be used to calculate the appropriate sample size, the estimated sample size was calculated using GPower 3.1 and EpiInfo. At a power of 95%, α margin of error of 5%, and effect size of 30%, a sample of 580 participants was needed to demonstrate statistical differences of appropriate power.

Ethical Considerations

This study was approved by the University of Jordan Scientific Committee.

Results

Demographics

A total of 1201 participants fully completed the questionnaire. Participants were characterized by a mean age of 36.3 (SD 14.4) years and a male-to-female ratio of nearly 1:1. Most participants had at least a university degree (ie, bachelor's; 758/1201, 63.1%), lived in the capital of Jordan (743/1201, 61.9%), and resided within urban cities (1056/1201, 87.9%). Nearly 58% (690/1201, 57.5%) of all participants were married, of which 70.7% (488/1201) had at least 3 children. Furthermore, 28.6% (344/1201) held office-based occupations, 26.2% (315/1201) were unemployed, 26.0% (312/1201) were students, 19.2% (230/1201) held field occupations, 4.4% (53, 1201) were housewives, and 3.9% (47/1201) were retired.

In terms of electronic competency, 54.3% (652/1201) perceived that they were highly competent in dealing with computers or electronic tablets. The most common internet connection types reported were 4G standard routers (386/1201, 30.6%), followed by fiber optic internet (522/1201, 23.5%). The sociodemographic characteristics of the cohort are described in Table 1.

Table 1. Demographics of the recruited participants (n=1201).

Variable	Results, n (%)
Gender	
Male	615 (51.2)
Female	586 (48.8)
Age (years)	
18-30	515 (42.9)
31-40	232 (19.3)
41-50	227 (18.9)
51-60	165 (13.7)
61-85	62 (5.2)
Educational level	
Primary	29 (2.4)
High school	209 (17.4)
University	758 (63.1)
Postgraduate (ie, master's)	205 (17.1)
Area	
Capital (Amman)	743 (61.9)
Central	210 (17.5)
North	198 (16.5)
South	50 (4.2)
Residence	
Urban	1056 (87.9)
Rural	145 (12.1)
Relationship status	
Married	690 (57.5)
Single	511 (42.5)
Number of children	
0	57 (8.3)
1	51 (7.4)
2	94 (13.6)
≥3	488 (70.7)
Nature of work	
Office	344 (28.6)
Field	230 (19.2)
Retired	47 (3.9)
Student	312 (26.0)
Housewife	53 (4.4)
Unemployed	215 (17.9)
Type of internet connection	
Fiber	522 (23.5)
4G router	368 (30.6)
ADSL	45 (3.7)
Phone packets	184 (15.3)

Variable	Results, n (%)
Router and phone packets	81 (6.7)
None	1 (0.1)
Comorbidities (yes)	280 (23.3)
Medications (yes)	359 (29.9)
Perceived ability	
High (8-10)	652 (54.3)
Low (0-7)	549 (45.7)

Knowledge of Telemedicine

Table 2 shows that 51.5% (619/1201) had heard of the concept of telemedicine, while only 24.9% (299/1201) had observed it in action. Also, only a mere 14.2% (170/1201) had tried using telemedicine within their lifetime. Nonetheless, participants demonstrated favorable views regarding telemedicine, as the greater majority agreed that telemedicine reduces the number of medical staff (755/1201, 62.9%), reduces transportation costs and time (1127/1201, 93.8%), facilitates better health care for older adults (948/1201, 78.9%), allows you to remotely conduct follow-up health care logistics (eg, acquiring a prescription refill; 1018/1201, 84.8%), and enables remote and close follow-up of patients' illnesses (824/1201, 68.6%).

Table 2. Participants' responses on the telemedicine knowledge items.

Knowledge items	Agree, n (%)	Disagree, n (%)
Have you heard of telemedicine before?	619 (51.5)	582 (48.5)
Have you observed a telemedicine process previously?	299 (24.9)	902 (75.1)
Have you ever used telemedicine before?	170 (14.2)	1031 (85.8)
Telemedicine reduces the number of needed medical staff.	755 (62.9)	446 (37.1)
Telemedicine reduces transportation costs and time.	1127 (93.8)	74 (6.2)
Telemedicine facilitates care for older adult patients.	948 (78.9)	253 (21.1)
With telemedicine, you can remotely acquire a prescription, refill, or admission orders	1018 (84.8)	183 (15.2)
With telemedicine, you can closely follow up with your ongoing/chronic illnesses remotely.	824 (68.6)	377 (31.4)

Attitudes Toward Telemedicine

With regards to participants' attitudes, our cohort displayed mostly favorable attitudes toward telemedicine. The majority of participants believed that telemedicine is useful during a pandemic (1001/1201, 83.3%), is able to decrease the number of outpatient visits (1033/1201, 86.0%), can increase the speed of performing health care services (893/1201, 74.4%), mitigates health care costs (897/1201, 74.7%), and is able to provide specialized health care to underserved areas (841/1201, 70.0%). Overall, 67.8% (814/1201) of all participants were willing to use telemedicine for diagnosis or follow up. Table 3 displays participants' attitudes and perceptions.

On the other hand, a significant number of participants believed that telemedicine cannot provide accurate diagnoses (810/1201, 67.4%) or comprehensive health care (757/1201, 63.0%). Moreover, the greater majority of participants believed that telemedicine is unable to reduce medical errors (889/1201, 74.0%). Most importantly, 43.5% (522/1201) of all participants believed that telemedicine may pose a threat to their information privacy. In terms of perceptions, most participants believed that Jordan does not have the capacity to implement telemedicine (626/1201, 60.4%); however, more than one-half (654/1201, 54.5%) of the cohort perceived telemedicine as the future of health care.

Table 3. Attitudes and perceptions of respondents toward telemedicine.

Attitudes and perceptions	Response categories, n (%)				Score, mean (SD)
	Strongly disagree	Disagree	Agree	Strongly agree	
Attitudes					
Is telemedicine useful during a pandemic (eg, COVID-19)?	71 (5.9)	129 (10.7)	401 (33.4)	600 (50.0)	3.3 (0.9)
Can telemedicine accurately facilitate the diagnosis of people?	332 (27.6)	478 (39.8)	278 (23.1)	113 (9.4)	2.1 (0.9)
Does telemedicine improve the communication between patient and physician?	177 (14.7)	325 (27.1)	420 (35.0)	279 (23.2)	2.7 (0.9)
Does telemedicine decrease visits to outpatient clinics?	69 (5.7)	99 (8.2)	454 (37.8)	579 (48.2)	3.3 (0.8)
Does telemedicine help increase the speed of performing medical care?	105 (8.7)	203 (16.9)	492 (41.0)	401 (33.4)	2.9 (0.9)
Does telemedicine reduce medical errors?	415 (34.6)	474 (39.5)	204 (17.0)	108 (9.0)	2.0 (0.9)
Is telemedicine able to provide patients with comprehensive health care?	271 (22.5)	486 (40.5)	302 (25.1)	142 (11.8)	2.3 (0.9)
Does telemedicine threaten information privacy?	270 (22.5)	409 (34.1)	331 (27.6)	191 (15.9)	2.4 (1.0)
Does telemedicine reduce the costs of providing health care?	87 (7.2)	217 (18.1)	510 (42.5)	387 (32.2)	2.9 (0.9)
Are you willing to use telemedicine for your medical diagnosis or follow-up?	189 (15.7)	198 (16.5)	456 (38.0)	358 (29.8)	2.8 (1.0)
Telemedicine will improve access to specialized health care for people who live in rural and suburban areas.	153 (12.7)	207 (17.2)	455 (37.9)	386 (32.1)	2.9 (0.9)
Perceptions					
Do you believe that Jordan has the capacity to adopt and implement telemedicine services?	338 (28.1)	288 (32.3)	306 (25.5)	169 (14.1)	2.3 (1.0)
Telemedicine can be integrated within the existing system.	177 (14.7)	319 (26.6)	457 (38.1)	248 (20.6)	2.6 (0.9)
Telemedicine is the future of clinical practice.	187 (15.6)	360 (30.0)	427 (35.6)	227 (18.9)	2.6 (0.9)

Factors Associated With Knowledge and Attitudes Toward Telemedicine

Univariate analysis demonstrated that having higher educational degrees (P<.001), living in urban districts (P=.002), and having a higher perception of electronics usage ability (P<.001) were associated with higher knowledge of telemedicine. Similarly, having a higher educational level (P=.006), residing within the capital (P<.001), having a fast internet connection (P=.004), and having a higher perception of electronics usage ability (P<.001) were associated with more positive attitudes toward telemedicine. Furthermore, female gender was associated with more positive perceptions of the implementation of telemedicine within Jordan (P=.007). Table 4 delineates the associations between attitudes, knowledge, perceptions, and different sociodemographic variables.

On another note, it appears that being married (odds ratio [OR] 1.426, 95% CI 1.017 to 1.998; P=.04), having children (OR 1.498, 95% CI 1.074 to 2.089; P=.02), having comorbidities (OR 1.770, 95% CI 1.244 to 2.517; P=.002), and taking medications (OR 1.695, 95% CI 1.212 to 2.371; P=.003) were associated with a higher likelihood of telemedicine usage.

Table 4. Differences in knowledge, attitude, and perception scores across different sociodemographic variables.

Variables	Knowledge score, mean (SD)	*P* value	Attitudes score, mean (SD)	*P* value	Perception score, mean (SD)	*P* value
Gender						
Male	4.8 (1.5)	.96	30.2 (6.6)	.25	7.3 (2.5)	.007
Female	4.8 (1.5)		29.7 (5.9)		7.6 (2.2)	
Age (years)						
18-30	4.8 (1.5)	.41	30.3 (5.7)	.05	7.5 (2.3)	.80
31-40	4.7 (1.6)		29.7 (6.5)		7.4 (2.4)	
41-50	4.9 (1.6)		29.8 (6.5)		7.4 (2.4)	
51-60	4.8 (1.6)		30.0 (7.1)		7.4 (2.6)	
61-85	4.6 (1.5)		27.8 (6.5)		7.2 (2.5)	
Educational level						
Primary	4.4 (1.4)	<.001	29.0 (6.6)	.006	7.8 (3.0)	.39
High school	4.3 (1.5)		28.9 (7.5)		7.4 (2.7)	
University	4.8 (1.5)		29.9 (5.8)		7.4 (2.3)	
Postgraduate (ie, masters)	5.2 (1.4)		31.1 (6.5)		7.7 (2.3)	
Area						
Capital (Amman)	4.8 (1.5)	.23	30.1 (3.1)	<.001	7.5 (2.4)	<.001
Central	4.7 (1.3)		29.7 (6.7)		7.4 (2.3)	
North	4.8 (1.6)		30.6 (6.3)		7.7 (2.2)	
South	4.4 (2.3)		27.1 (7.4)		6.1 (2.6)	
Residence						
Urban	4.8 (1.5)	.002	30.1 (3.2)	.14	7.5 (2.3)	.05
Rural	4.4 (1.7)		29.2 (7.1)		7.1 (2.6)	
Relationship status						
Married	4.8 (1.6)	.74	29.7 (6.7)	.09	7.5 (2.3)	.36
Single	4.8 (1.5)		30.3 (5.7)		7.4 (2.5)	
Number of children						
0	4.7 (1.5)	.83	30.1 (6.0)	.64	7.6 (2.3)	.38
1	4.8 (1.7)		30.5 (6.9)		7.4 (2.4)	
2	4.8 (1.6)		29.5 (6.3)		7.2 (2.1)	
≥3	4.8 (1.6)		29.7 (6.6)		7.4 (2.5)	
Type of internet connection						
4G	4.7 (1.5)	.11	30.2 (6.3)	.004	7.5 (2.3)	.72
Fiber	4.9 (1.5)		30.4 (3.2)		7.5 (2.4)	
Phone packets, others	4.7 (1.7)		28.9 (6.5)		7.4 (2.5)	
Perceived ability						
High (8-10)	4.9 (1.5)	<.001	30.7 (6.0)	<.001	7.6 (2.3)	.11
Low (0-7)	4.6 (1.6)		29.0 (6.5)		7.4 (2.4)	

Predictors of Positive Attitudes Toward Telemedicine

The multivariate linear regression analysis demonstrated that the perceived ability of using electronics was associated with positive attitudes (β=0.394, 95% CI 0.224 to 0.563), while living in Southern Jordan predicted poor attitudes toward telemedicine (β=–2.896, 95% CI –4.873 to –0.919). Table 5 provides the results of the multivariate regression model. Type, and therefore speed, of internet connection did not influence attitudes levels toward telemedicine.

Table 5. Factors predicting favorable attitudes toward telemedicine.

Factors	Linear regression model for attitudes			
	P value	B	β	Lower and upper 95% CIs for B
Age	.74	–0.007	0.022	–0.051 to 0.036
Gender (female)	.59	–0.199	0.369	–0.922 to 0.525
Number of children	.83	0.026	0.122	–0.214 to 0.266
Ability to deal with computers and tablets	<.001	0.394	0.086	0.224 to 0.563
Presence of comorbidities	.87	–0.110	0.671	–1.427 to 1.207
Taking regular medications	.64	0.294	0.620	–0.922 to 1.510
Being Married	.95	–0.034	0.612	–1.234 to 1.165
High school education	.49	–0.871	1.272	–3.367 to 1.624
University education	.57	–0.699	1.243	–3.137 to 1.740
Postgraduate education	.88	0.188	1.304	–2.370 to 2.746
4G internet	.08	0.859	0.495	–0.111 to 1.829
Fiber internet	.07	0.897	0.491	–0.067 to 1.861
Living in Amman	.08	–0.963	0.546	–2.035 to 0.109
Living in Central Jordan	.09	–1.064	0.629	–2.298 to 0.170
Living in Southern Jordan	.004	–2.896	1.008	–4.873 to –0.919

Discussion

Principal Findings

In summary, we demonstrated that about one-half of our studied population were aware of the existence of telemedicine, while only a small minority declared having observed or ever used the technology. Furthermore, more than one-half of the population were willing to use telemedicine. Overall, attitudes toward telemedicine were positive. However, participants questioned its capacity to reduce errors, make accurate diagnoses, or provide comprehensive care. Moreover, telemedicine was perceived as a threat to privacy and impractical within Jordan's current resource capacity. Attitudes and degree of knowledge were typically higher in those with higher education levels, who perceived themselves as competent electronics users, and who lived in metropolitan regions. Regression analysis showed that perceived competency with electronics was a positive predictor of favorable attitudes, while living in rural areas (Southern Jordan) was a negative predictor. Also, a higher likelihood of using telemedicine was associated with having children, being married, having previous comorbidities, or taking medications.

To the best of our knowledge, this study is the first of its kind in Jordan. We aimed to explore the levels of awareness, knowledge, and concerns of the general public toward telemedicine in an attempt to closely inspect the understanding of its potential recipients and providers of the concept of the technology and further evaluate the possibility of incorporating it into the current health system. Within the regional literature, studies conducted in Saudi Arabia and Egyptian populations showcased similar results to those of our study [30,31]. Similar results were also found among citizens of the United Arab Emirates [17]. The aforementioned populations had positive perceptions of telemedicine, its benefits, and its applications within the health care system. However, significant proportions of both populations agreed, to an extent, that telemedicine poses a threat to patients' privacy and may increase medical errors. Throughout the literature, especially studies conducted in response to the pandemic, it appears that the utility of telemedicine during disasters, its ability to reduce transportation time and costs, and its capacity to improve patient-doctor communication are consistently appreciated [30]. On the other hand, older Asian adults (ie, Chinese, Malaysian, and Indian participants) reported negative perceptions toward the utility and usefulness of telemedicine [34]. Nonetheless, despite the general positive perception of telemedicine, the literature describes that the most commonly reported individual hurdles to successful implementation of telemedicine technology worldwide were related to lack of awareness [35-37] and technical literacy [38,39]. Tackling those factors early through promotion of telemedicine services and enhancing electronic use would likely facilitate smooth integration of the technology into the current system.

In terms of attitudes, our participants demonstrated positive perceptions toward telemedicine. They believed that it has the capacity to decrease transportation costs and the number of medical staff required. They also believed that telemedicine can smoothen health care logistics and facilitate better health care for older adults. Mutual economic benefits for patients and health care providers were evidently recognized in successful telemedicine projects in Jordan [21]. Furthermore, an integrative review [40] displayed evidence supporting the facilitation of delivered care and improved health outcomes for older adults through the media of telemedicine. Such a value is particularly essential in the era of COVID-19 as well as among older adults, who possess a higher risk of infection and complications from

the disease and could benefit from reducing hospital visits whenever possible [41]. This is further relevant to Jordan's medical landscape, as it is characterized by patient congestion and long waiting times at nearly every public medical institution. These factors could increase the risk of infection for older adults, thus positively affecting the perceptions of the caretakers and relatives of older adults toward the usefulness of telemedicine.

Moreover, Jordanian participants believed that telemedicine helps in reducing outpatient visits, increasing the speed of delivering health care services, and providing more comprehensive care to underserved regions. Also, they perceive it as a useful tool during a pandemic similar to that of COVID-19. Similar statements were also observed from residents of Saudi Arabia and Egypt [30,31]. In conjunction with social distancing and work-from-home policies, telemedicine represented a viable and sustainable option to migrate health care from hospitals to homes, which promoted reductions in outpatient clinic visits and further aided in the endeavors to mitigate COVID-19 transmission [6]. In addition, resource-limited countries often experience a scarcity in health care providers, particularly specialists. Therefore, telemedicine could also provide a novel approach to improve access to more specialized wellness programs for vast majorities of people as well as revitalizing weakened public health services while significantly reducing costs [42].

On the other hand, our study participants expressed a multitude of concerns toward telemedicine. These concerns were mostly pertaining to accuracy of diagnosis, medical errors, and information privacy. Diagnostic accuracy through telemedicine services has been thoroughly evaluated in the literature. Most studies reported high accuracy in specialties where telemedicine is viable, although variations across literature are still evident. Most notably, those specialties were dermatology [43] and ophthalmology [44,45].

Although some studies associate the rapid expansion of telemedicine technology in COVID-19 with an increased incidence of medical errors [46], a root cause analysis of medical errors done by the Agency for Healthcare Research and Quality Patient Safety Network [47] revealed that most incidents were related to human factors such as communication failures between health care personnel. Telemedicine, when implemented properly and utilized by competent personnel, could serve as a tool to help mitigate those errors by providing continuous surveillance to delivered care, especially in critical settings such as that of the intensive care unit [48].

Maintaining privacy and information security was a challenging conundrum that was threatened and prominently highlighted through the era of rapid expansion of telemedicine services during COVID-19. Such challenges were manifested as a lack of controls or limits on the collection, use, and disclosure of sensitive personal information, repeated cyberattacks, and rapid spread of ransomware and went as far as suspected death in patients [49]. Such hazards should be dealt with by a multidisciplinary approach including health care organizers, policy makers, and information technology specialists and by the implication of guidelines and rules that regulate the use of telemedicine services similar to those of the Health Insurance Portability and Accountability Act regulating traditional medicine. Efforts should be concentrated at training both the health care provider and the patient about the necessary precautions to prevent any breach of confidentiality and at enforcing end-to-end data encryption of patients' information and records in order to mitigate those adverse effects and sustain a culture of security [50].

In our study, higher rates of telemedicine use were significantly higher among patients with comorbidities and those taking chronic medications. Such results could be largely explained by the fact that, in 2020, due to COVID-19 restrictions, the Ministry of Health, along with the Royal Medical Services and university hospitals, offered primarily medication refill services through telemedicine programs [51]. This largely promoted ease of accessibility and was a good substitute to in-person appointments.

Among the studied cohort, participants who were married and had children had a higher likelihood of telemedicine usage. The particular use of telemedicine services within the context of a marriage appears to be a solution for a myriad of obstacles that otherwise would prevent couples from seeking therapy such as childcare, scheduling difficulties, and stigma. Providing couples treatment via telemedicine services was reported to address these obstacles and improve accessibility to health care services [52,53]. Additionally, the use of telemedicine services to perform regular prenatal counseling was associated with high rates of feasibility and satisfaction [54]; such factors could encourage multiparous women who might be more lenient when it comes to prenatal visits to have a positive perspective regarding telemedicine within this context, to save time and effort.

More than one-half of the studied population believed that Jordan cannot successfully integrate telemedicine into the current health care system. This may be due to resource scarcity, as it remains a major hindrance for the implementation of telemedicine in Jordan. Considering its position as a country with limited resources and a high poverty rate of 24.1% [55], the implementation of telemedicine in an ideal manner in which high satisfaction and wide coverage of various medical specialties for the majority of the population are achieved remains a struggle that necessitates urgent policy changes, making the development of the health care sector as a whole of upmost priority in the country in order to facilitate the path for optimal effectuation of telemedicine.

Another reason for such negative perceptions of telemedicine is the misleading status of the technological infrastructure in Jordan. Although around 67% of the Jordanian population utilize the internet [56], internet quality and speed might be the greatest hindrance to proper implementation of telemedicine, particularly in remote areas. The cheapest home internet subscription in Jordan costs around US $50, and, considering that the average minimum wage is around US $370, people from a lower socioeconomic status will be discouraged from having the necessary resources to benefit from telemedicine [57]. It should be noted that only 33% of Jordanians owned a personal computer or laptop as of 2017 [58]. Interestingly, our prediction model did not correlate the type of internet with attitudes toward

telemedicine. This may be a result of the survey distribution process, as it was primarily online, thus targeting participants with better internet connections.

Limitations

Our results should be considered with caution due to the following limitations: the cross-sectional study design and its implications, close-ended nature of the questionnaire that may miss certain participants' responses, and sampling technique that may have missed certain groups of the Jordanian populace (eg, older adults and technologically illiterate individuals).

Conclusions

In light of these findings, Jordanians have favorable and positive perceptions of telemedicine that are similar to other populations within the same region. Nonetheless, concerns with regards to privacy, medical errors, and capacity for making accurate diagnoses are prevalent. Despite the generally positive perceptions toward and willingness to use telemedicine, Jordanians believed that telemedicine cannot be easily integrated within the current system nor do they have resources to adopt it.

Conflicts of Interest

None declared.

References

1. Telemedicine: Opportunities and developments in Member States. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/44497 [accessed 2022-10-25]
2. Wittson CL, Benschoter R. Two-way television: helping the medical center reach out. Am J Psychiatry 1972 Nov;129(5):624-627. [doi: 10.1176/ajp.129.5.624] [Medline: 4673018]
3. Fryer K, Delgado A, Foti T, Reid CN, Marshall J. Implementation of obstetric telehealth during COVID-19 and beyond. Matern Child Health J 2020 Sep;24(9):1104-1110 [FREE Full text] [doi: 10.1007/s10995-020-02967-7] [Medline: 32564248]
4. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. Ir J Med Sci 2021 Feb 08;190(1):1-10 [FREE Full text] [doi: 10.1007/s11845-020-02299-z] [Medline: 32642981]
5. Li P, Liu X, Mason E, Hu G, Zhou Y, Li W, et al. How telemedicine integrated into China's anti-COVID-19 strategies: case from a National Referral Center. BMJ Health Care Inform 2020 Aug;27(3):1 [FREE Full text] [doi: 10.1136/bmjhci-2020-100164] [Medline: 32830110]
6. Okereafor K, Adebola O, Djehaiche R. Exploring the potentials of telemedicine and other non-contact electronic health technologies in controlling the spread of the novel coronavirus disease (COVID-19). International Journal in IT & Engineering (IJITE) 2020:1-13. [doi: 10.6084/m9.figshare.12479111.v1]
7. Ronis SD, McConnochie KM, Wang H, Wood NE. Urban telemedicine enables equity in access to acute illness care. Telemed J E Health 2017 Feb;23(2):105-112. [doi: 10.1089/tmj.2016.0098] [Medline: 27383822]
8. Committee on Pediatric Workforce, Marcin JP, Rimsza ME, Moskowitz WB. The use of telemedicine to address access and physician workforce shortages. Pediatrics 2015 Jul;136(1):202-209. [doi: 10.1542/peds.2015-1253] [Medline: 26122802]
9. Kuroda N. Decision making on telemedicine for patients with epilepsy during the coronavirus disease 2019 (COVID-19) Crisis. Front Neurol 2020 Jun 26;11:722 [FREE Full text] [doi: 10.3389/fneur.2020.00722] [Medline: 32714276]
10. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
11. Adler G, Pritchett LR, Kauth MR, Nadorff D. A pilot project to improve access to telepsychotherapy at rural clinics. Telemed J E Health 2014 Jan;20(1):83-85 [FREE Full text] [doi: 10.1089/tmj.2013.0085] [Medline: 24168721]
12. Petersen C, DeMuro P. Legal and regulatory considerations associated with use of patient-generated health data from social media and mobile health (mHealth) devices. Appl Clin Inform 2017 Dec 19;06(01):16-26. [doi: 10.4338/aci-2014-09-r-0082]
13. Jalghoum YA, Khasawneh SS. An empirical research of the challenges to e-health initiative in Jordan. International Journal of Business and Social Science 2016;7(11):51-65 [FREE Full text]
14. Alaboudi A, Atkins A, Sharp B, Alzahrani M, Balkhair A, Sunbul T. Perceptions and attitudes of clinical staff towards telemedicine acceptance in Saudi Arabia. 2017 Presented at: IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA); November 29 - December 2, 2016; Agadir, Morocco. [doi: 10.1109/aiccsa.2016.7945714]
15. Alkabba A, Hussein GA, Albar A, Bahnassy A, Qadi M. The major medical ethical challenges facing the public and healthcare providers in Saudi Arabia. J Family Community Med 2012 Jan;19(1):1-6 [FREE Full text] [doi: 10.4103/2230-8229.94003] [Medline: 22518351]
16. Albarrak AI, Mohammed R, Almarshoud N, Almujalli L, Aljaeed R, Altuwaijiri S, et al. Assessment of physician's knowledge, perception and willingness of telemedicine in Riyadh region, Saudi Arabia. J Infect Public Health 2021 Jan;14(1):97-102 [FREE Full text] [doi: 10.1016/j.jiph.2019.04.006] [Medline: 31060975]
17. Abdool S, Abdallah S, Akhlaq S, Razzak HA. User acceptance level of and attitudes towards telemedicine in the United Arab Emirates: a quantitative study. Sultan Qaboos Univ Med J 2021 May;21(2):e203-e209 [FREE Full text] [doi: 10.18295/squmj.2021.21.02.008] [Medline: 34221467]

18. Qaddoumi I, Nawaiseh I, Mehyar M, Razzouk B, Haik BG, Kharma S, et al. Team management, twinning, and telemedicine in retinoblastoma: a 3-tier approach implemented in the first eye salvage program in Jordan. Pediatr Blood Cancer 2008 Aug 25;51(2):241-244. [doi: 10.1002/pbc.21489] [Medline: 18300320]
19. El Hayek S, Nofal M, Abdelrahman D, Adra A, Al Harthi M, Al Shamli S, et al. Telepsychiatry in the Arab world: a viewpoint before and during COVID-19. NDT 2020 Nov;Volume 16:2805-2815. [doi: 10.2147/ndt.s277224]
20. Odeh R, Gharaibeh L, Daher A, Kussad S, Alassaf A. Caring for a child with type 1 diabetes during COVID-19 lockdown in a developing country: Challenges and parents' perspectives on the use of telemedicine. Diabetes Res Clin Pract 2020 Oct;168:108393 [FREE Full text] [doi: 10.1016/j.diabres.2020.108393] [Medline: 32858098]
21. Shdaifat MBM, Khasawneh RA, Alefan Q. Clinical and economic impact of telemedicine in the management of pediatric asthma in Jordan: a pharmacist-led intervention. J Asthma 2022 Jul 08;59(7):1452-1462. [doi: 10.1080/02770903.2021.1924774] [Medline: 33941032]
22. Khader Y, Jarrah M, Al-Shudifat A, Shdaifat A, Aljanabi H, Al-Fakeh S, et al. Telecardiology application in jordan: its impact on diagnosis and disease management, patients' quality of life, and time- and cost-savings. Int J Telemed Appl 2014;2014:819837 [FREE Full text] [doi: 10.1155/2014/819837] [Medline: 25400661]
23. Hakeem Electronic Health Solutions. URL: https://emed.hakeem.jo/index-en.html [accessed 2022-10-25]
24. King Hussein Cancer Center Chooses tebcan for Online Consultations with Patients. Tebcan Tebline. 2020 Jul 22. URL: https://tebcan.com/en/Jordan/tebline/Admin/Oncology/Articles/King-Hussein-Cancer-Center-Choses-tebcan-for-Online-Consultations-with-Patients_3119 [accessed 2022-10-25]
25. Yousef Y, Al-Nawaiseh I, Mehyar M, Sultan I, Al-Hussaini M, Jaradat I, et al. How telemedicine and centralized care changed the natural history of retinoblastoma in a developing country: analysis of 478 patients. Ophthalmology 2021 Jan;128(1):130-137. [doi: 10.1016/j.ophtha.2020.07.026] [Medline: 32682835]
26. Telehealth. Abdali Hospital. URL: https://www.abdalihospital.com/telehealth [accessed 2022-10-25]
27. About. Altibbi. URL: https://altibbi.com/%D8%B9%D9%86-%D8%A7%D9%84%D8%B7%D8%A8%D9%8A [accessed 2022-10-25]
28. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
29. Elsaie ML, Shehata HA, Hanafi NS, Ibrahim SM, Ibrahim HS, Abdelmaksoud A. Egyptian dermatologists attitude toward telemedicine amidst the COVID19 pandemic: a cross-sectional study. J Dermatolog Treat 2022 Mar;33(2):1067-1073. [doi: 10.1080/09546634.2020.1800576] [Medline: 32723123]
30. Alajwari H, Alfayez A, Alsalman D, Alanezi F, Alhodaib H, Al-Rayes S, et al. Knowledge and attitude of Saudi Arabian citizens towards telemedicine during the COVID-19 pandemic. Int Health 2021 Dec 03:6 [FREE Full text] [doi: 10.1093/inthealth/ihab082] [Medline: 34893850]
31. Alboraie M, Allam M, Youssef N, Abdalgaber M, El-Raey F, Abdeen N, et al. Knowledge, applicability, and barriers of telemedicine in Egypt: a national survey. Int J Telemed Appl 2021;2021:5565652 [FREE Full text] [doi: 10.1155/2021/5565652] [Medline: 34211550]
32. Zwingelberg SB, Mercieca K, Elksne E, Scheffler S, Prokosch V. [Attitude of patients to possible telemedicine in ophthalmology : Survey by questionnaire in patients with glaucoma]. Ophthalmologe 2022 Apr;119(4):374-380 [FREE Full text] [doi: 10.1007/s00347-021-01501-6] [Medline: 34542691]
33. Ahmed MH, Awol SM, Kanfe SG, Hailegebreal S, Debele GR, Dube GN, et al. Willingness to use telemedicine during COVID-19 among health professionals in a low income country. Inform Med Unlocked 2021;27:100783 [FREE Full text] [doi: 10.1016/j.imu.2021.100783] [Medline: 34778509]
34. Teo CL, Chee ML, Koh KH, Tseng RMWW, Majithia S, Thakur S, et al. COVID-19 awareness, knowledge and perception towards digital health in an urban multi-ethnic Asian population. Sci Rep 2021 May 24;11(1):10795 [FREE Full text] [doi: 10.1038/s41598-021-90098-6] [Medline: 34031469]
35. Rutledge CM, Haney T, Bordelon M, Renaud M, Fowler C. Telehealth: preparing advanced practice nurses to address healthcare needs in rural and underserved populations. Int J Nurs Educ Scholarsh 2014 Jan 14;11:1. [doi: 10.1515/ijnes-2013-0061] [Medline: 24423469]
36. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
37. Ross J, Stevenson F, Lau R, Murray E. Exploring the challenges of implementing e-health: a protocol for an update of a systematic review of reviews. BMJ Open 2015 Apr 13;5(4):e006773 [FREE Full text] [doi: 10.1136/bmjopen-2014-006773] [Medline: 25869685]
38. Yu J, Afridi SM, Cozart AC, Isea L, Guan J. Evaluation and feedback for telehealth from patients and physicians during the early stage of COVID-19 pandemic period. Cureus 2021 Jan 11;13(1):e12633 [FREE Full text] [doi: 10.7759/cureus.12633] [Medline: 33585121]

39. Alkire (née Nasr) L, O'Connor GE, Myrden S, Köcher S. Patient experience in the digital age: An investigation into the effect of generational cohorts. Journal of Retailing and Consumer Services 2020 Nov;57:102221. [doi: 10.1016/j.jretconser.2020.102221]
40. Groom LL, McCarthy MM, Stimpfel AW, Brody AA. Telemedicine and telehealth in nursing homes: an integrative review. J Am Med Dir Assoc 2021 Sep;22(9):1784-1801.e7. [doi: 10.1016/j.jamda.2021.02.037] [Medline: 33819450]
41. COVID-19 Risks and Vaccine Information for Older Adults. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/aging/covid19/covid19-older-adults.html [accessed 2022-10-25]
42. Scott R, Mars M. Telehealth in the developing world: current status and future prospects. SHTT 2015 Feb:25. [doi: 10.2147/SHTT.S75184]
43. Giavina-Bianchi M, Giavina-Bianchi P, Santos AP, Rizzo LV, Cordioli E. Accuracy and efficiency of telemedicine in atopic dermatitis. JAAD Int 2020 Dec;1(2):175-181 [FREE Full text] [doi: 10.1016/j.jdin.2020.08.002] [Medline: 34409337]
44. Ullah W, Pathan SK, Panchal A, Anandan S, Saleem K, Sattar Y, et al. Cost-effectiveness and diagnostic accuracy of telemedicine in macular disease and diabetic retinopathy: A systematic review and meta-analysis. Medicine (Baltimore) 2020 Jun 19;99(25):e20306 [FREE Full text] [doi: 10.1097/MD.0000000000020306] [Medline: 32569163]
45. Kang S, Dehabadi M, Sim DA, Thomas PBM, Appiah Ewusi D, Ezra D. Accuracy of periocular lesion assessment using telemedicine. BMJ Health Care Inform 2021 Jan;28(1):1 [FREE Full text] [doi: 10.1136/bmjhci-2020-100287] [Medline: 33483345]
46. Bashshur RL, Bashshur MJ, Krupinski EA. Telemedicine, precision medicine, and regionalization. Telemed J E Health 2022 May 09;28(5):599-601 [FREE Full text] [doi: 10.1089/tmj.2021.0357] [Medline: 34515531]
47. Root Cause Analysis. Agency for Healthcare Research and Quality Patient Safety Network. 2019 Sep 07. URL: https://psnet.ahrq.gov/primers/primer/10/Root-Cause-Analysis [accessed 2022-10-25]
48. Becker CD, Dandy K, Gaujean M, Fusaro M, Scurlock C. Legal perspectives on telemedicine part 2: telemedicine in the intensive care unit and medicolegal risk. Perm J 2019;23:1 [FREE Full text] [doi: 10.7812/TPP/18.294] [Medline: 31496501]
49. Eddy M, Perlroth N. Cyber Attack Suspected in German Woman's Death. The New York Times. 2020 Sep 18. URL: https://www.nytimes.com/2020/09/18/world/europe/cyber-attack-germany-ransomeware-death.html [accessed 2022-10-25]
50. Jalali MS, Landman A, Gordon WJ. Telemedicine, privacy, and information security in the age of COVID-19. J Am Med Inform Assoc 2021 Mar 01;28(3):671-672 [FREE Full text] [doi: 10.1093/jamia/ocaa310] [Medline: 33325533]
51. Obeidat AZ, El-Salem K. A national telemedicine program in the Kingdom of Jordan - Editorial. Ann Med Surg (Lond) 2021 Feb;62:145-149 [FREE Full text] [doi: 10.1016/j.amsu.2021.01.009] [Medline: 33520212]
52. Hogan JN. Conducting couple therapy via telehealth: special considerations for virtual success. J Health Serv Psychol 2022;48(2):89-96 [FREE Full text] [doi: 10.1007/s42843-022-00060-x] [Medline: 35496919]
53. Orlowski EW, Friedlander ML, Megale A, Peterson EK, Anderson SR. Couple and family therapists' experiences with Telehealth during the COVID-19 pandemic: a phenomenological analysis. Contemp Fam Ther 2022;44(2):101-114 [FREE Full text] [doi: 10.1007/s10591-022-09640-x] [Medline: 35400810]
54. Oelmeier K, Schmitz R, Möllers M, Braun J, Deharde D, Sourouni M, et al. Satisfaction with and feasibility of prenatal counseling via telemedicine: a prospective cohort study. Telemed J E Health 2022 Aug;28(8):1193-1198. [doi: 10.1089/tmj.2021.0309] [Medline: 34861131]
55. Weldali M. Poverty rate of 24.1% requires policy change — experts. The Jordan Times. 2022 Jun 09. URL: http://www.jordantimes.com/news/local/poverty-rate-241-requires-policy-change-%E2%80%94-experts [accessed 2022-10-25]
56. Jordan. The World Bank - Data. URL: https://data.worldbank.org/country/JO [accessed 2022-10-25]
57. Gaith B. Minimum wage falls short of living costs — experts. The Jordan Times. 2022 Jul 31. URL: https://jordantimes.com/news/local/minimum-wage-falls-short-living-costs-%E2%80%94-experts [accessed 2022-10-25]
58. T J. ICT, Internet usage among Jordanian households increases — ministry. The Jordan Times. 2018 Dec 26. URL: https://jordantimes.com/news/local/ict-internet-usage-among-jordanian-households-increases-%E2%80%94-ministry [accessed 2022-10-25]

Abbreviations

OR: odds ratio

Edited by A Kushniruk; submitted 27.07.22; peer-reviewed by P Codyre, W Cheng, X Liu; comments to author 30.09.22; revised version received 03.10.22; accepted 06.10.22; published 04.11.22.

Please cite as:
Murshidi R, Hammouri M, Taha H, Kitaneh R, Alshneikat M, Al-Qawasmeh A, Al-Oleimat A, Al-Huneidy L, Al-Huneidy Y, Al-Ani A Knowledge, Attitudes, and Perceptions of Jordanians Toward Adopting and Using Telemedicine: National Cross-sectional Study JMIR Hum Factors 2022;9(4):e41499
URL: https://humanfactors.jmir.org/2022/4/e41499
doi:10.2196/41499
PMID:36257609

Original Paper

Using QR Codes as a Form of eHealth to Promote Health Among Women in a Pandemic: Cross-sectional Study

Natalia Fischer-Suárez[1*], MD, PhD; David Lozano-Paniagua[2*], PhD; Sonia García-Duarte[1*], CNM; Gracia Castro-Luna[2], MD, PhD; Tesifón Parrón-Carreño[2*], MD, PhD; Bruno José Nievas-Soriano[2*], MD, PhD

[1]Obstetrics and Gynecology Unit, Torrecárdenas Hospital, Almeria, Spain

[2]Nursing Science, Physiotherapy and Medicine Department, University of Almeria, Almeria, Spain

[*]these authors contributed equally

Corresponding Author:
David Lozano-Paniagua, PhD
Nursing Science, Physiotherapy and Medicine Department
University of Almeria
Ctra Sacramento, s/n
Almeria, 04120
Spain
Phone: 34 950 21 45 69
Email: dlozano@ual.es

Abstract

Background: QR codes have played an integral role during the pandemic in many sectors, but their use has been limited in the health care sector, especially by patients. Although some authors have stated that developing specific content for women on how to cope with health problems could be an effective way to prevent problems, especially during pandemics, there is little research regarding the use of QR codes to promote health during a pandemic, and even fewer studies are focused on women. Moreover, although the importance of assessing these interventions from the users' perspective has been stated, research carried out from this point of view is still scarce.

Objective: This study aimed to assess the usefulness of using QR codes with information to promote women's health in the context of a pandemic. We also sought to design and validate a questionnaire to assess this.

Methods: A cross-sectional study was conducted among women in the gynecology waiting rooms of a reference hospital. Exploratory factorial analysis with the split-half method and Cronbach α values was performed for questionnaire validation. Univariant and bivariant analyses were performed to analyze the data obtained.

Results: In total, 186 women took part in the study. Exploratory factor analysis identified 2 domains: usability and applicability in medical practice. The Cronbach α value was .81. Overall, 83.7% of the answers to the first domain and 56.4% of those to the second were favorable. Women with university education or those who had used QR codes before scored better in the usability domain, while no differences were observed in the applicability scores.

Conclusions: Using QR codes in the gynecology clinics' waiting rooms can help promote women's health during a pandemic, regardless of their education level or whether they have used QR codes before. The questionnaire developed herein is a helpful tool to assess this. These findings are important for clinical practice. This research can be performed in other ambits, specialties, or countries.

(JMIR Hum Factors 2022;9(4):e41143) doi:10.2196/41143

KEYWORDS

eHealth; QR code; women; health promotion; public health; healthcare sector; women's health; gynecology; gynecologist; gynecological; obstetrician; obstetric; healthcare; health care; pandemic; questionnaire; validation; validate; development; cross sectional; factorial analysis

Introduction

COVID-19, an infectious disease caused by SARS-CoV-2, was declared a global pandemic on March 11, 2020 [1]. This pandemic has been a turning point for health promotion [2]. Health promotion aims to enable people to increase control over their health. Health promotion may be more crucial than ever during a pandemic and can contribute to addressing the threat at different levels [2]. However, many measures to prevent citizens and health care personnel from becoming infected have involved a change in behavior [2]. Social interactions have decreased, reducing the number of medical consultations [3].

During the COVID-19 pandemic, there has been a growing interest in using technology in delivering health care services to users as an infection control measure owing to the removal of brochures and other health information in waiting rooms. QR codes are accessible tools that can be created effortlessly and freely by health care professionals. They can be later scanned through several smartphone apps or even using the camera app in many smartphones such as the newer iPhones [4]. QR codes are not a new concept [1]. QR codes allow the development of simple, cost-effective, and functional systems based on the optical recognition of inexpensive barcodes attached to physical objects. Combined with website platforms, these systems can provide helpful services that are already broadly used in many other contexts [5]. As other authors have described, using these codes is easy and straightforward [6,7].

This technology, integrated with face-to-face communication, may improve shared decision-making [7], as QR codes are very convenient for the average user, and their launching among professionals and final users is relatively straightforward [5].

QR codes have played an integral role during the pandemic in many sectors, but their use has been limited in the health care sector, especially by patients [1]. A possible explanation is that although health care QR code systems are promising, they are not without problems [8], especially those concerning security when using QR codes to transmit personal information [9]. Although some authors have stated that developing specific content for women on how to cope with health problems could be an effective way to prevent problems, especially during pandemics [10], there is little research regarding the use of QR codes to promote health during a pandemic, and even fewer studies focused on women. Moreover, although the importance of assessing these interventions from the users' perspective has been stated [11,12], research carried out from this point of view is still scarce [13,14].

Thus, we wanted to assess a new tool in a distinct population, which had not been considered before in a specific environment. Therefore, the main aim of this research was to assess the usefulness of QR codes with information to promote women's health in the context of a pandemic. The specific objectives of this study are to design a questionnaire with the least number of items possible to evaluate the usefulness of QR codes with information to promote health, to validate the questionnaire through construct validation and reliability tests, and to apply the validated questionnaire to evaluate the usefulness of the use of QR codes with information to promote health in women attending gynecology and obstetrics consultations.

Methods

Study Design

A cross-sectional observational study was conducted within the context of the COVID-19 pandemic to evaluate the usefulness of QR codes with information to promote women's health during the pandemic. The QR codes were available in the gynecology and obstetrics outpatient clinics' waiting rooms at the Torrecardenas University Hospital. This hospital is the reference for the province of Almeria, Spain, located in the southeast of Spain, with a population size of 731,792 inhabitants as of January 1, 2022, according to the Spanish National Statistics Institute [15].

The questionnaire was developed from scratch on the basis of data obtained from a review of the literature and the authors' experience. The questionnaire collected 5 demographic aspects (age, educational level, the reason for consultation, whether QR codes had been used previously, and the topic consulted through the available QR codes). It also collected 8 qualitative items that assessed the experience of using QR codes and the perceived usefulness in providing health information for women. Likert scales were used, with answers that ranged from "nothing" to "a lot."

Sample Size Estimation

The sample size was estimated using the Epi Info app (Centers for Disease Control and Prevention, Atlanta, Georgia, United States) with the following parameters: population size 358,656 (female population in the province of Almería, as of January 1, 2022) [15]; 80% CI, and a level of precision of 5%. These parameters revealed a required sample size of 164 participants. The authors determined to collect at least 180 answers to diminish potential auto-selection bias—this value was in accordance with the guideline specified by Kline et al [16] of using 2-20 subjects for each questionnaire item for the factorial analysis.

Eligible Population and Recruitment

The eligible population comprised women aged 18 or older who attended gynecology and obstetrics outpatient clinics and were offered to participate. The inclusion criteria were as follows: being women, 18 or older, speaking and understanding Spanish, and being recruited in the waiting room of gynecology and obstetrics clinics. The exclusion criteria were as follows: not meeting any of the referred inclusion criteria and not wishing to participate in the study, despite meeting the criteria.

Questionnaire Validation

The adequacy of the exploratory factor analysis was determined using the Bartlett test and the Kaiser-Meyer-Olkin measure. To evaluate construct validity, the 8 qualitative items of the questionnaire were assessed through exploratory factor analysis. The tool's reliability was measured using Cronbach α [17]. The split-half method was used to assess stability because the questionnaire could not be retested with the same users [18].

Statistical Analyses

Statistical analyses were conducted using SPSS (version 28; IBM Corp). Univariant and bivariant analyses were performed. The scoring of each domain was the sum of the scores provided by the participants when answering the questions. These scores could range from 1 to 4 points in each question. Thus, the score range for the first domain was 4-16 and that for second domain was 2-8. The scores of each domain were used to perform bivariate analyses regarding different aspects of the participants. Informed consent was shown at the beginning of the questionnaire. Personal data were not collected.

Ethics Approval

This study was approved by the Research and Ethics Committee of Nursing, Physiotherapy, and Medicine Department of the University of Almeria, Spain, (EFM 200/2020). The questionnaire did not collect personal information.

Results

Sociodemographic Results

In total, 186 women took part in the research, with completed questionnaires with valid responses. Their mean age was 34.3 (SD 7.5) years. In total, 124 (67.7%) participants had nonuniversity studies (Table 1), and 171 (91.9%) were waiting for consultations regarding pregnancy, assisted reproduction, general gynecology, and emergencies. In total, 136 (73.1%) had used QR codes before.

Among the QR codes available in the waiting room, the most consulted ones were pregnancy physical activity, contraceptive methods, and pregnancy nutrition, for 149 (80.1%) women (Table 2). The least consulted topic was gender violence, consulted by 4 (2.1%) women.

Table 1. Sociodemographic characteristics of the participants (N=186).

Characteristics	Values
Age (years), mean (SD)	34.3 (7.5)
Education level, n (%)	
Primary	13 (7.0)
Secondary	58 (31.2)
Other	35 (18.8)
Baccalaureate	18 (9.7)
University	62 (33.3)
Medical consultation, n (%)	
Pregnancy	67 (36.0)
Assisted reproduction	43 (23.1)
General gynecology	39 (21.0)
Emergencies	22 (11.8)
Other	7 (3.8)
Lower genital tract	3 (1.6)
Breast	2 (1.1)
Oncology	2 (1.1)
Pelvic floor	1 (0.5)
Had used QR codes before, n (%)	
Yes	136 (73.1)
No	50 (26.9)
Total	186 (100.0)

Table 2. Topic consulted using available QR codes.

Topic	Participants, n (%)
Pregnancy physical activity	74 (39.8)
Contraceptive methods	45 (24.2)
Pregnancy nutrition	30 (16.1)
Breastfeeding	12 (6.5)
Emergency contraception	13 (7.0)
Vital wills	8 (4.3)
Gender violence	4 (2.1)
Total	186 (100.0)

Validation of the Questionnaire

The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.779, and the Bartlett test for sphericity was 368.6 (P<.001). These results indicate the model's suitability for exploratory factor analysis, which excluded 2 of the 8 items from the questionnaire, identifying 2 domains that explained 68.6% of the variance (Table 3). The first domain was defined by 4 items; the second one, by 2 items (Table 4).

These domains were interpreted by analyzing the items within each domain (Textbox 1). Thus, the first domain defined usability and the second one defined applicability in medical practice.

The Cronbach α of the questionnaire was .81. The Cronbach α value for the first domain was .77, while that of the second domain was .80. The split-half method did not reveal significant differences in the domains or the global evaluation (Table 5).

Table 3. Total explained variance determined through principal component analysis.

Component	Total	Percentage of variance, %	Accumulated percentage, %
1	3.099	51.7	51.7
2	1.018	16.9	68.6
3	0.682	11.4	—[a]
4	0.474	7.9	—
5	0.412	6.9	—
6	0.315	5.2	—

[a]Not available.

Table 4. Rotated component matrix. Extraction method: principal component analysis; rotation method: varimax with Kaiser normalization.

Item	Component	
	1	2
Q1	0.688	—[a]
Q3	0.778	—
Q6	—	0.876
Q7	—	0.881
Q8	0.575	—
Q2	0.862	—

[a]Not available.

Textbox 1. Domains detected.

Domain 1 : usability
Q1. Do you think that the information received through this initiative can help you in your daily life?
Q3. Did you find the information provided easy to understand?
Q8. Do you think this initiative could be helpful in other medical specialties?
Q2. Did you find it convenient to use QR codes to access information?
Domain 2 : applicability in medical practice
Q6. Has the information allowed you to clarify doubts before entering the medical consultation?
Q7. Has the information helped you to enter the medical consultation more informed?

Table 5. Split-half method.

Domain	Participants, n	Score, mean (SD)	*P* value[a]
Patients' usability			.19
First half	93	10.7 (2.7)	
Second half	93	10.8 (2.3)	
Applicability in medical practice			.90
First half	93	5.4 (2.3)	
Second half	93	5.0 (2.2)	
Total			.77
First half	93	16.1 (4.4)	
Second half	93	15.8 (4.0)	

[a]Mann-Whitney test.

Univariant Analysis

Regarding the evaluation of the domains defined by the questionnaire (Table 6), 83.7% of the answers to the questions that defined the first domain (usability) and 56.4% of the answers to the questions of the second domain (applicability in medical practice) were favorable (responses being "something," "quite," and "a lot").

Table 6. Answers of the participants (N=186).

Questions	Participants, n (%)
Domain 1: usability	
Q1. Do you think the information received can help you in your daily life?	
Nothing	11 (5.9)
A bit	20 (10.8)
Something	67 (36.0)
Quite	68 (36.6)
A lot	20 (10.8)
Q3. Did you find the information provided easy to understand?	
Nothing	12 (6.5)
A bit	13 (7.0)
Something	39 (21.0)
Quite	82 (44.1)
A lot	40 (21.5)
Q8. Do you think this initiative could be helpful in other medical specialties?	
Nothing	9 (4.8)
A bit	16 (8.6)
Something	41 (22.0)
Quite	70 (37.6)
A lot	50 (26.9)
Q2. Did you find it convenient to use QR codes to access information?	
Nothing	18 (9.7)
A bit	17 (9.1)
Something	48 (25.8)
Quite	70 (37.6)
A lot	33 (17.7)
Domain 2: applicability in medical practice	
Q6. Has the information allowed you to clarify doubts before entering the consultation?	
Nothing	49 (26.3)
A bit	35 (18.8)
Something	67 (36.0)
Quite	21 (11.3)
A lot	14 (7.5)
Q7. Has the information helped you to enter the medical consultation more informed?	
Nothing	50 (26.9)
A bit	28 (15.1)
Something	63 (33.9)
Quite	32 (17.2)
A lot	13 (7.0)

Bivariant Analysis

When analyzing the score of the domains for different aspects, no differences were found regarding the age of the participants in the first domain (P=.13, Spearman ρ) or in the second domain (P=.25, Spearman ρ). Regarding education level (Figure 1), statistically significant differences were found in the scores of the first domain, where the participants with university education

scored higher than those with primary school education (*P*=.01, Kruskal-Wallis test for independent samples). No differences were found with regard to education level when analyzing the second domain (*P*=.10, Kruskal-Wallis test for independent samples).

When analyzing the score of both domains regarding the medical consultations of the patients (Figure 2), no differences were found (first domain, *P*=.55; second domain, *P*=.44; Kruskal-Wallis test for independent samples).

When analyzing the score of both domains regarding if the participants had used QR codes before (Figure 3), those who had used QR codes before scored higher in the first domain (*P*=.003, Mann-Whitney *U* test). However, no differences were found in the scores of the second domain regarding previous use of QR codes (*P*=.13, Mann-Whitney *U* test).

When analyzing the score of the domains regarding the topic consulted using the available QR codes (Figure 4), higher scores of the first domain were achieved by women who consulted the *Breastfeeding*, *Pregnancy Nutrition*, and *Emergency Contraception* topics (*P*<.001, Kruskal-Wallis test for independent samples). In the second domain, higher scores were achieved by women who consulted the *Breastfeeding* and *Pregnancy Nutrition* topics (*P*<.001, Kruskal-Wallis test for independent samples).

Figure 1. Domains' scores regarding the education level of participants.

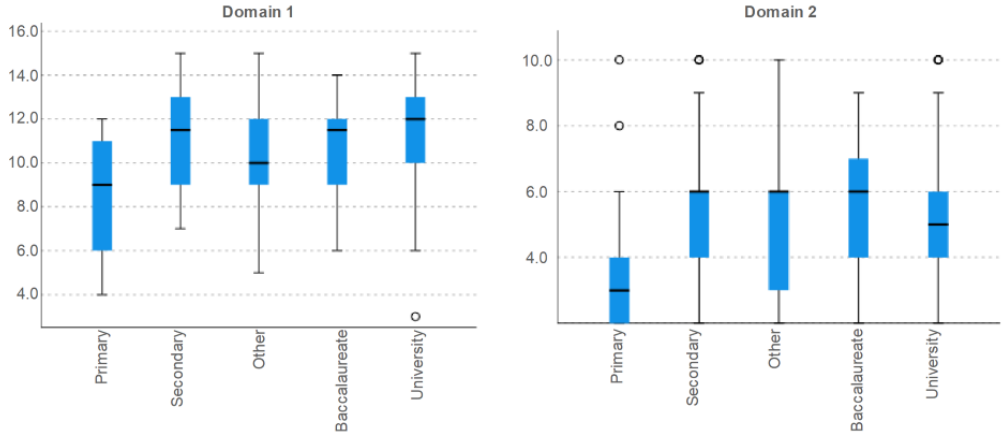


Figure 2. Domains' scores regarding medical consultation of participants.

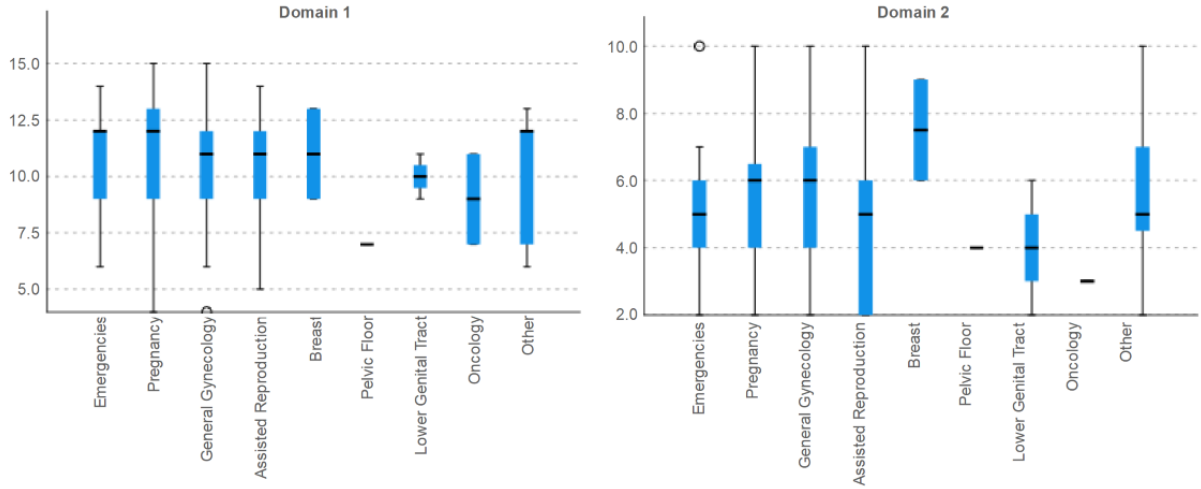


Figure 3. Domains' scores regarding participants' previous use of QR codes.

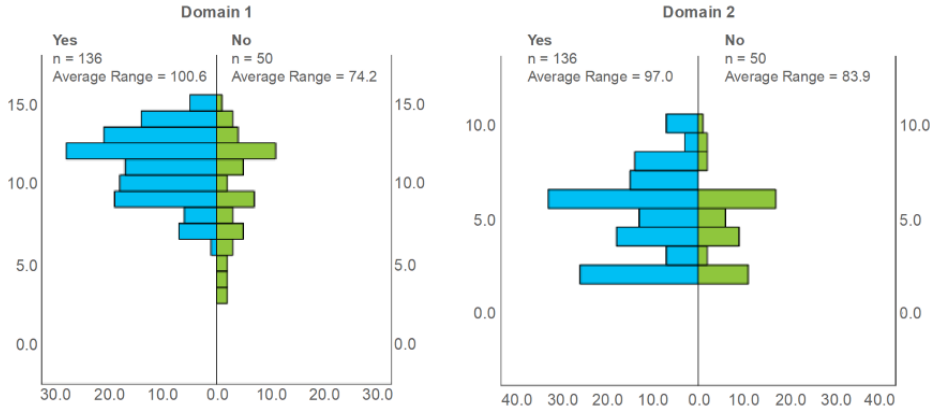
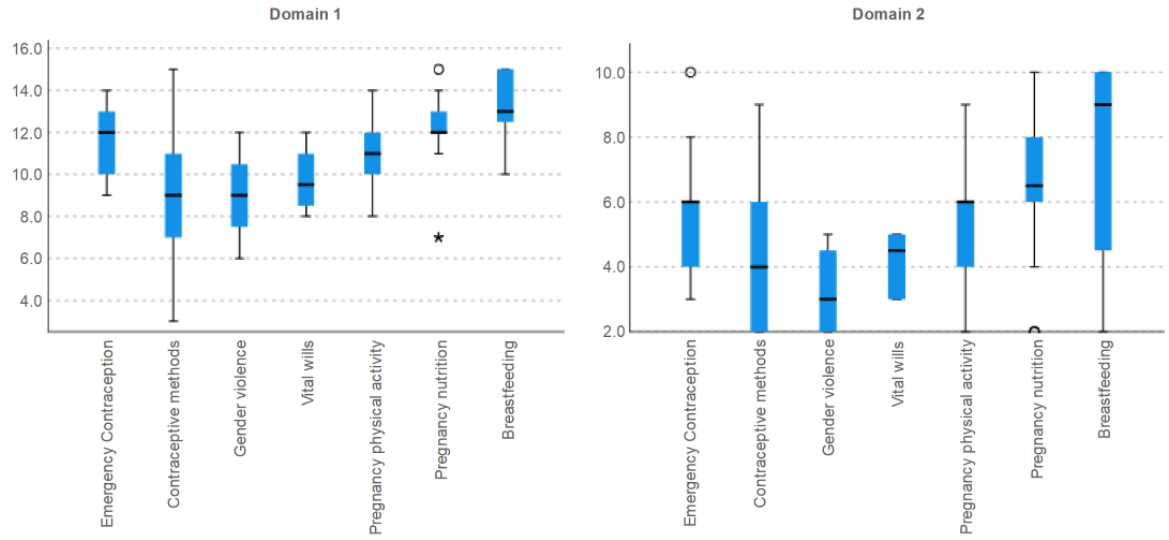


Figure 4. Domains' scores regarding topic consulted.



Discussion

Principal Findings

Sociodemographic Findings

According to some studies [19-21], women who use eHealth resources often have higher education levels. However, in our study, most of the participants had nonuniversity education. This finding is logical as the recruitment process was carried out in the waiting room during hospital gynecology consultations, where the eligible population is more representative of the women of the general population than of women who often use eHealth resources. We believe that this aspect adds value to the research, given that the participant women were more representative of the general population, contrary to many eHealth-focused studies [21]. In these studies, women tend to participate to a greater extent as they are more prone to using eHealth, committing auto-selection bias, and having more education.

Validation of the Questionnaire

The exploratory factor analysis excluded 2 items from the initial questionnaire, depicting 2 domains assessed on the basis of the items that defined them, as reported previously [22]. Although the Q8 item factor loading value was 0.575, which may seem low, this value was clearly greater than those for the discarded items, it was very close to 0.6, and it allowed us to obtain a higher Cronbach α for the questionnaire. The Cronbach α coefficient is the most used method to evaluate the internal consistency of the questionnaire [22], and the value obtained by the final questionnaire can be considered good or even excellent, according to previous studies [22,23]. The split-half method, applied in the same period or when other methods including test-retest cannot be used [24], confirmed the stability of the questionnaire.

Univariant Analysis

Some studies have stated that, regarding the provision of health care services to users, QR codes are accessible, free, easy to use, and can be scanned through several free smartphone apps [4]. However, few studies have assessed the usability and applicability of these QR codes among patients in medical consultations. Even fewer studies have evaluated these aspects specifically in women and in the context of a pandemic.

Regarding usability from the perspective of the women in this study, similar research in other eHealth ambits and general users has stated that simple designs and previous testing of the interventions with the patients can help ensure an overall good perception of usability [25-27]. The items that defined this domain in our resultant questionnaire regarded aspects that have

been previously described separately by other authors, such as the easiness of understanding, convenience of using QR codes to access the information [28], and helpfulness in daily life [29] or in other medical specialties [1]. Therefore, we can consider that these items can help define the usability of the QR codes and the health information provided through them from the perspective of general women. Most participants provided positive responses to the items that defined this domain; hence, we can conclude that their perception of the usability of the QR codes for health promotion was good or very good.

Regarding applicability in medical practice, other authors have stated that QR codes can help avoid unnecessary medical consultations [30] and allow patients to access health information or their medical records [31]. In our study, the items that defined this domain regarded clarifying doubts and offering information related to medical consultation—aspects that have been stated in other studies [28]. Thus, these aspects can help assess the QR codes' applicability in medical consultations. More than half of the participants responded positively to the items that defined this domain. Considering the diversity of our sample, less used to eHealth interventions than the ones referred to in other studies [12,32,33], where the use of convenience samples is common [33], we can conclude that the perception of the women regarding the applicability of the QR codes in the medical consultation was also good.

Bivariant Analysis

Similar to previous studies [12,34], we found that the age of the participants did not influence the score of the domains. This finding is essential, as it is usually described that young adults are more prone to using QR codes, especially in health aspects [35]. Another interesting finding is that the first domain received better scores from the women with university studies when compared to those with nonuniversity studies. As reported previously [1], it is understandable that women with university students found it easier to use QR codes and reading the information offered.

Moreover, no differences were found in the second domain, which defined the applicability regarding education level. This finding implies that every user, regardless of their education level, could benefit from using QR codes for health promotion. It is also important that when analyzing the evaluations of women regarding the medical consultations they attended, no differences were found in any of the domains. Another logical finding is that the women who had used QR codes assigned better scores to the first domain: usability. However, no differences were found in the second domain—applicability—regarding this same aspect. Again, this is an essential finding because every woman could benefit from using QR codes for health promotion regardless of whether they had used them before.

In summary, the usability of QR codes to promote health aspects among women during a pandemic is better perceived by women with university education and those who have used QR codes before. However, more importantly, the applicability of the information offered through the QR codes is independent of their age, education level, the medical consultation attended, or whether they had previously used QR codes. In other words, their applicability can be considered similar for every woman.

Limitations and Strengths

This research has some limitations. The most important is the selection bias due to various factors. Our sample was obtained from the waiting room of the gynecology clinic of a reference hospital, but it may not be representative of other regions or countries. Women who did not speak or read in Spanish or were younger than 18 years were excluded. We must also consider that the questionnaire and the study were in Spanish. Thus, the final validated questionnaire could be applied in other hospitals or translated and culturally adapted to other languages or countries for future research. Although the sample size in our study was larger than that in another similar study [36], larger sample sizes could also help increase the CI. These aspects must be considered when assessing the external validity of our conclusions. Finally, the potential security issues of familiarizing people to scan random signs in the waiting rooms of medical practices should be considered. For instance, malicious individuals could quite easily place signs in public places. While this is beyond the scope of this study, it is an important consideration that needs to be considered when the use of QR codes.

This study also has some strengths. The most important one is that the questionnaire yielded favorable results in the exploratory factor analysis and the value of Cronbach α. These aspects provide validity to the final questionnaire and the results obtained. Another important strength is that the study has been performed with actual patients from the waiting rooms of gynecology clinics. Considering the demographic characteristics of the sample, it seems to be more representative of actual women who can benefit from using QR codes to improve their health in real life. Thus, our findings can be helpful for current clinical practice and future research.

Conclusions

The questionnaire developed here has been demonstrated to be a helpful tool to measure the perception of the usability of QR codes and the applicability of the information offered even before entering the medical consultations. Thus, considering the answers provided by the participants, we can conclude that using QR codes in the consultations' waiting rooms can help promote women's health during a pandemic, regardless of their age, education level, or whether they have previously used QR codes. These last findings are essential because the applicability can be similar for every woman. These findings are crucial for current clinical practice. Finally, this research can be adapted to be performed in other ambits, specialties, or even different countries, and compare the results obtained, so the focus of this method can be widened and extended.

Conflicts of Interest

None declared.

References

1. Sharara S, Radia S. Quick Response (QR) codes for patient information delivery: A digital innovation during the coronavirus pandemic. J Orthod 2022 Mar;49(1):89-97. [doi: 10.1177/14653125211031568] [Medline: 34308694]
2. Van den Broucke S. Why health promotion matters to the COVID-19 pandemic, and vice versa. Health Promot Int 2020 Apr 01;35(2):181-186 [FREE Full text] [doi: 10.1093/heapro/daaa042] [Medline: 32297931]
3. Wang J, McDonald N, Cochran AL, Oluyede L, Wolfe M, Prunkl L. Health care visits during the COVID-19 pandemic: a spatial and temporal analysis of mobile device data. Health Place 2021 Nov;72:102679 [FREE Full text] [doi: 10.1016/j.healthplace.2021.102679] [Medline: 34628150]
4. Wahsheh H, Al-Zahrani M. Secure and usable QR codes for healthcare systems: the case of Covid-19 pandemic. 2021 Presented at: 12th International Conference on Information and Communication Systems (ICICS); May 24-26, 2021; Valencia. [doi: 10.1109/icics52457.2021.9464565]
5. Sanz-Valero J, Álvarez Sabucedo LM, Wanden-Berghe C, Santos Gago JM. QR codes: outlook for food science and nutrition. Crit Rev Food Sci Nutr 2016;56(6):973-978. [doi: 10.1080/10408398.2012.742865] [Medline: 26047907]
6. Masih EA. Nurse Educ Today 2022 Apr;111:105305 [FREE Full text] [doi: 10.1016/j.nedt.2022.105305] [Medline: 35220149]
7. Hsu P, Wu C, Kuo L, Chen M, Chen Y, Huang S, et al. Improving the process of shared decision-making by integrating online structured information and self-assessment tools. J Pers Med 2022 Feb 10;12(2):256 [FREE Full text] [doi: 10.3390/jpm12020256] [Medline: 35207744]
8. Wang T, Jia F. The impact of health QR code system on older people in China during the COVID-19 outbreak. Age Ageing 2021 Jan 08;50(1):55-56 [FREE Full text] [doi: 10.1093/ageing/afaa222] [Medline: 32990312]
9. Lin P, Lan W, Chen Y, Wu W. A confidential QR code approach with higher information privacy. Entropy (Basel) 2022 Feb 16;24(2):284 [FREE Full text] [doi: 10.3390/e24020284] [Medline: 35205577]
10. Jiang H, Jin L, Qian X, Xiong X, La X, Chen W, et al. Maternal mental health status and approaches for accessing antenatal care information during the COVID-19 epidemic in China: cross-sectional study. J Med Internet Res 2021 Jan 18;23(1):e18722 [FREE Full text] [doi: 10.2196/18722] [Medline: 33347423]
11. Linardon J, Messer M, Lee S, Rosato J. Perspectives of e-health interventions for treating and preventing eating disorders: descriptive study of perceived advantages and barriers, help-seeking intentions, and preferred functionality. Eat Weight Disord 2021 May;26(4):1097-1109. [doi: 10.1007/s40519-020-01005-3] [Medline: 32959274]
12. Nievas-Soriano BJ, García-Duarte S, Fernández-Alonso AM, Bonillo-Perales A, Parrón-Carreño T. Users evaluation of a Spanish eHealth pediatric website. Comput Methods Programs Biomed 2021 Nov;212:106462 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106462] [Medline: 34715515]
13. Li X. Understanding eHealth literacy from a privacy perspective: eHealth literacy and digital privacy skills in American disadvantaged communities. American Behavioral Scientist 2018 Jul 09;62(10):1431-1449 [FREE Full text] [doi: 10.1177/0002764218787019]
14. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psychiatry 2018 Oct 18;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
15. Instituto Nacional de Estadística. URL: https://www.ine.es/index.htm [accessed 2022-07-14]
16. Kline R. Principles and Practice of Structural Equation Modeling (fourth Edition). New York, NY: Gilford Press; 2016.
17. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
18. Boparai JK, Singh S, Kathuria P. How to design and validate a questionnaire: a guide. Curr Clin Pharmacol 2018;13(4):210-215. [doi: 10.2174/1574884713666180807151328] [Medline: 30084336]
19. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
20. Villadsen S, Hadi H, Ismail I, Osborne R, Ekstrøm CT, Kayser L. ehealth literacy and health literacy among immigrants and their descendants compared with women of Danish origin: a cross-sectional study using a multidimensional approach among pregnant women. BMJ Open 2020 May 07;10(5):e037076 [FREE Full text] [doi: 10.1136/bmjopen-2020-037076] [Medline: 32385065]
21. Nievas Soriano BJ, García Duarte S, Fernández Alonso AM, Bonillo Perales A, Parrón Carreño T. Validation of a questionnaire developed to evaluate a pediatric eHealth website for parents. Int J Environ Res Public Health 2020 Apr 13;17(8):2671 [FREE Full text] [doi: 10.3390/ijerph17082671] [Medline: 32295030]

22. Paiva PCP, de Paiva HN, de Oliveira Filho PM, Lamounier JA, Ferreira e Ferreira E, Ferreira RC, et al. Development and validation of a social capital questionnaire for adolescent students (SCQ-AS). PLoS One 2014 Aug 5;9(8):e103785 [FREE Full text] [doi: 10.1371/journal.pone.0103785] [Medline: 25093409]
23. Castro R, Oliveira SD, Pereira IDO, Santos WD, Fernandes S, Silva RAR. Construct validation: coping with HIV/AIDS in primary health care. Rev Bras Enferm 2019 Sep 16;72(5):1173-1181 [FREE Full text] [doi: 10.1590/0034-7167-2018-0734] [Medline: 31531638]
24. Sürücü L, Maslakçi A. Validity and reliability in quantitative research. Bus Manag Stud 2020 Sep 25;8(3):2694-2726. [doi: 10.15295/bmij.v8i3.1540]
25. Zhou L, DeAlmeida D, Parmanto B. Applying a user-centered approach to building a mobile personal health record app: development and usability study. JMIR Mhealth Uhealth 2019 Jul 05;7(7):e13194 [FREE Full text] [doi: 10.2196/13194] [Medline: 31278732]
26. Doyle S, Pavlos R, Carlson S, Barton K, Bhuiyan M, Boeing B, et al. Efficacy of digital health tools for a pediatric patient registry: semistructured interviews and interface usability testing with parents and clinicians. JMIR Form Res 2022 Jan 17;6(1):e29889 [FREE Full text] [doi: 10.2196/29889] [Medline: 35037889]
27. Htet K, Phyu A, Thwin T, Chongsuvivatwong V. Mobile health app for tuberculosis screening and compliance to undergo chest x-ray examination among presumptive cases detected by the app in Myanmar: usability study. JMIR Form Res 2022 Jun 07;6(6):e37779 [FREE Full text] [doi: 10.2196/37779] [Medline: 35623000]
28. Karia C, Hughes A, Carr S. Uses of quick response codes in healthcare education: a scoping review. BMC Med Educ 2019 Dec 06;19(1):456 [FREE Full text] [doi: 10.1186/s12909-019-1876-4] [Medline: 31810464]
29. Brodie K, Madden L, Rosen C. Applications of quick response (QR) codes in medical education. J Grad Med Educ 2020;12(2):138-140. [doi: 10.4300/jgme-d-19-00516.1]
30. Faggiano A, Carugo S. Can the implementation of electronic surveys with quick response (QR) codes be useful in the COVID-19 era? Int J Epidemiol 2020 Oct 01;49(5):1732-1733 [FREE Full text] [doi: 10.1093/ije/dyaa170] [Medline: 33097946]
31. Banjar HR, Zaher GF, Almutiry HS, Alshamarni ASA, Almouhana GI, Alahwal HM, et al. Web-based expert system with quick response code for beta-thalassemia management. Health Informatics J 2021 Feb 11;27(1):1460458221989397 [FREE Full text] [doi: 10.1177/1460458221989397] [Medline: 33570008]
32. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. BMJ 2004 May 15;328(7449):1150 [FREE Full text] [doi: 10.1136/bmj.328.7449.1150] [Medline: 15142895]
33. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
34. Schomakers E, Lidynia C, Ziefle M. Listen to my heart? How privacy concerns shape users' acceptance of e-Health technologies. 2019 Presented at: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); October 21-23, 2019; Barcelona. [doi: 10.1109/wimob.2019.8923448]
35. Tai Z, Yu X, He B. Locked down through virtual disconnect: navigating life by staying on/off the health QR code during COVID-19 in China. Convergence 2021 Sep 30;27(6):1648-1662 [FREE Full text] [doi: 10.1177/13548565211047157]
36. Çaka SY, Çınar N. Effect of education on the readiness levels of primigravida women to hygienic care practices of the newborn with the use of QR code. Health Care Women Int 2021 May 03:1-17 [FREE Full text] [doi: 10.1080/07399332.2021.1908289] [Medline: 33939938]

Edited by G Eysenbach; submitted 17.07.22; peer-reviewed by N Baharom, K Heffernan; comments to author 16.08.22; revised version received 02.09.22; accepted 18.10.22; published 08.11.22.

Please cite as:
Fischer-Suárez N, Lozano-Paniagua D, García-Duarte S, Castro-Luna G, Parrón-Carreño T, Nievas-Soriano BJ
Using QR Codes as a Form of eHealth to Promote Health Among Women in a Pandemic: Cross-sectional Study
JMIR Hum Factors 2022;9(4):e41143
URL: https://humanfactors.jmir.org/2022/4/e41143
doi:10.2196/41143
PMID:36346657

Original Paper

Behavior Change App for Self-management of Gestational Diabetes: Design and Evaluation of Desirable Features

Mikko Kytö[1,2], MSc, PhD; Saila Koivusalo[3,4,5,6], MD, PhD; Antti Ruonala[2], MA; Lisbeth Strömberg[2], MSocSc; Heli Tuomonen[2], BSc; Seppo Heinonen[3,4], MD, PhD; Giulio Jacucci[2], MSc, PhD

[1]Helsinki University Hospital IT Management, Helsinki University Hospital, Helsinki, Finland
[2]Department of Computer Science, University of Helsinki, Helsinki, Finland
[3]Department of Gynecology and Obstetrics, Turku University Hospital, Turku, Finland
[4]Department of Gynecology and Obstetrics, University of Turku, Turku, Finland
[5]Department of Gynecology and Obstetrics, Helsinki University Hospital, Helsinki, Finland
[6]Department of Gynecology and Obstetrics, University of Helsinki, Helsinki, Finland

Corresponding Author:
Mikko Kytö, MSc, PhD
Helsinki University Hospital IT Management
Helsinki University Hospital
Paciuksenkatu 25
Helsinki, 00270
Finland
Phone: 358 94711
Email: mikko.kyto@hus.fi

Abstract

Background: Gestational diabetes (GDM) has considerable and increasing health effects as it raises both the mother's and the offspring's risk for short- and long-term health problems. GDM can usually be treated with a healthier lifestyle, such as appropriate dietary modifications and sufficient physical activity. Although telemedicine interventions providing weekly or more frequent feedback from health care professionals have shown the potential to improve glycemic control among women with GDM, apps without extensive input from health care professionals are limited and have not been shown to be effective. Different features in personalization and support have been proposed to increase the efficacy of GDM apps, but the knowledge of how these features should be designed is lacking.

Objective: The aim of this study is to investigate how GDM apps should be designed, considering the desirable features based on the previous literature.

Methods: We designed an interactive GDM prototype app that provided example implementations of desirable features, such as providing automatic and personalized suggestions and social support through the app. Women with GDM explored the prototype and provided feedback in semistructured interviews.

Results: We identified that (1) self-tracking data in GDM apps should be extended with written feedback, (2) habits and goals should be highly customizable to be useful, (3) the app should have different functions to provide social support, and (4) health care professionals should be notified through the app if something unusual occurs. In addition, we found 2 additional themes. First, basic functionalities that are fast to learn by women with GDM who have recently received the diagnosis should be provided, but there should also be deeper features to maintain interest for women with GDM at a later stage of pregnancy. Second, as women with GDM may have feelings of guilt, the app should have a tolerance for and a supporting approach to unfavorable behavior.

Conclusions: The feedback on the GDM prototype app supported the need for desirable features and provided new insights into how these features should be incorporated into GDM apps. We expect that following the proposed designs and feedback will increase the efficacy of GDM self-management apps.

Trial Registration: ClinicalTrials.gov NCT03941652; https://clinicaltrials.gov/ct2/show/NCT03941652

(JMIR Hum Factors 2022;9(4):e36987) doi:10.2196/36987

KEYWORDS

gestational diabetes; mobile app; features; behavior change; digital health; eHealth; telehealth; self-tracking; self-management; personalized health care

Introduction

Approximately 14% of pregnant women are diagnosed with gestational diabetes mellitus (GDM) worldwide, with the highest prevalence in the Middle East and North Africa (27.6%) and in Southeast Asia (20.8%) and the lowest in Europe (7.8%) and North America and the Caribbean (7.1%) [1]. These recent prevalence findings are consistent with the high prevalence of overall diabetes in these regions, and it has been suggested that the same lifestyle risk factors that underly the overall diabetes, such as obesity and metabolic syndrome, also contribute to an increased risk of GDM. GDM is associated with a range of adverse short- and long-term consequences for both mother and child [2-4]. Although GDM is a temporary condition that lasts until the birth of the child, GDM increases the later risk of type 2 diabetes [5]. The primary treatment for GDM is through adjustments toward a heathier lifestyle, especially changing one's diet and increasing exercise [6,7]. It is critical that women with GDM be supported in this behavior change [8]. Although a large body of previous studies on supporting type 1 diabetes and type 2 diabetes self-management with apps exists (eg, [9]), these studies do not consider the design implications for diabetes self-management arising from temporality and pregnancy.

Telehealth interventions that provide frequent feedback from health care personnel on lifestyle and glucose levels have been shown to be effective in improving glycemic control among women with GDM [10]. However, providing this feedback to each woman with GDM by health care professionals daily requires a lot of human work and is costly. GDM management apps with less input from health care professionals are limited and have not been shown to be effective [11-13]. The knowledge on user experience with GDM apps is also limited, although recent studies have provided relevant findings [14-18]. According to a user study by Skar et al [17], the most important features of the GDM app researched were the overview of blood glucose levels, real-time feedback from fingertip measurements, and information about nutrition. However, their app did not provide any automatic feedback or social support, which Kytö et al [14] argued was among the desirable features of mobile apps for GDM self-management. Other important features of GDM self-management apps are related to competence to manage GDM, personalization, reliable information, support for dual processing (ie, habitual and sense-making reasoning), and integration to existing health care services [14].

Although the previous studies on desirable features [14] have considered *what* the GDM apps for behavior change should cover, *how* to design these features has remained largely open and feedback from women with GDM has not been considered. Therefore, the aim of this study is to gather the experience of women with GDM with implementations of the desirable features in a GDM app prototype.

Methods

Research Design

To investigate how the desirable features of GDM apps [14] should be incorporated, our research design consisted of 2 phases: First, we designed a GDM prototype app consisting of implementations of the features [14], and second, we evaluated the experiences of these features with semistructured interviews. The prototype was targeted toward women with GDM who manage their diabetes with lifestyle adjustments and do not require medication (eg, metformin or insulin). As we designed a prototype whose purpose was to gather experiences from a large scale of features, it was a horizontal, experience prototype [19]. Our focus was on the content of the different features, and their functionalities were not implemented in detail (ie, all the buttons and actions did not function).

The purpose of the interviews was to investigate how the women with GDM perceived the example implementations of the desirable features and how they perceived the app for supporting the management of GDM in general. Participants (n=10 women with GDM) were recruited from maternity and antenatal clinics in the Helsinki Metropolitan Area (Finland).

Ethical Approval

The study was performed in compliance with the Declaration of Helsinki and approved by the Ethics Committee of the Helsinki University Hospital (HUS-2165-2018-3).

GDM Prototype App

Design Rationale for the Prototype

The design of the prototype was based on previous work by Kytö et al [14], which provided initial suggestions on how desirable features could be designed (summarized in Table 1). We had a multidisciplinary research group consisting of health care personnel who are experienced in treating women with GDM (a gynecologist, midwives, and diabetes nurses), a social psychologist, a user researcher, and an interaction designer to brainstorm the implementations of these features in workshops. The final layout of the prototype was designed by an interaction designer.

Table 1. Desirable features of mobile apps supporting behavior change in management of GDM[a], as described by Kytö et al [14].

Feature	Detail
1. Increase competence to manage GDM with automatic feedback and interactive exploration.	This feature highlights the need for feedback on lifestyle and data exploration in the learning process to manage GDM. In apps, women with GDM have so far received feedback from health care professionals [10], and the perception and effectiveness of automated feedback are largely unknown [14]. Lentferink et al [20] showed that praise and suggestions are especially useful behavior change techniques in mobile apps supporting a healthier lifestyle.
2. Increase autonomy by enabling personalization.	The ways to manage GDM are highly individual and significantly vary from person to person [8]. Personalization has a great impact on the perceived effectiveness of telehealth care in women with GDM. However, it has so far been limited to personalization of the app but not to treatment suggestions [14].
3. Provide social support, especially from the partner.	Encouragement by people with close relationships is important for women with GDM [8,21-24]. The partner's support is seen as especially valuable in effecting a behavioral change in women with GDM, such as increasing exercise [25], but it has not been considered in existing GDM apps [14]. To support the partner's willingness to use a GDM app, Peyton et al [26] suggested that pregnancy apps should not be too feminine.
4. Support normal pregnancy and debunk myths about GDM.	Women with GDM want reliable information about their disease [21,27], and this information should be in the same app as information about pregnancy [17].
5. Support dual processing as pregnancy is life changing.	Both reflective and habitual processes should be supported in GDM apps. Especially, habits can be considered important with women diagnosed with GDM, as they are generally well motivated to manage GDM but overwhelmed in their daily life [8]. Motivation for maintaining a healthier lifestyle typically decreases after the birth of the baby [17]. Habits could help maintain a healthier lifestyle after the pregnancy also and thus decrease the probability of type 2 diabetes. Between reflective and habitual processes exists contextual fluid reasoning, which is used in creating simple rules between blood glucose levels and lifestyle [28].
6. Integrate the app with normal pregnancy and existing health care services.	In telehealth solutions for GDM management, tight integration between health professionals and women with GDM has been shown to be important [10]. This plays a significant role in the perceived usefulness of the GDM apps [29,30].

[a]GDM: gestational diabetes mellitus.

The Evaluated GDM Prototype App

The desirable features [14] were implemented in the prototype app (see Table 2, Figure 1, and Figure 2). The prototype was interactive in nature and ran on a mobile phone's browser. Users could tap buttons on the screen to navigate pages and scroll. The prototype was implemented using inVision [31], which creates browser-based prototypes. A browser is not the platform for the future actual app, as there will be native iOS and Android versions. In this study, participants were given an iPhone 6, which used the Safari browser to run the prototype.

Table 2. Implementations of desirable features [14] in the prototype app.

Feature	Implementation
1. Increase competence to manage GDM[a] with automatic feedback and interactive exploration.	We investigated this first desirable feature in the form of providing interactive self-tracking data exploration (Figure 1b) and feedback in the form of praise and suggestion (Figure 2a).
2. Increase autonomy by enabling personalization.	We sought to customize the app by providing 2 functionalities, (1) a habit tool (see Figure 1c) and (2) the ability to add a profile picture, name, and expected date of birth into the app.
3. Provide social support, especially from the partner.	We incorporated features to provide social support from both the partner and other women with GDM. For the partner, we designed a view where they could receive suggestions from the app (Figure 2a) and a "Shared habits" functionality that provided the possibility for the partners to do things together (Figure 2b). The content would be provided to the partner through a dedicated partner's version of the GDM app. For women with GDM, we designed a community challenge where they could support each other to achieve a common goal (Figure 2c).
4. Support normal pregnancy and debunk myths about GDM.	The app provided a draft information section about pregnancy (eg, development of the fetus, as shown in Figure 1d), GDM, nutrition, and physical activity.
5. Support dual processing as pregnancy is life changing.	We supported dual processing by implementing 2 different types of visualizations, an area graph to support reflective thinking (Figure 1b) and the latest values to support habitual thinking (Figure 1a). We also created a habit formation and tracking tool to support more autonomous behavior. The reflective visualization was incorporated by visualizing blood glucose levels, nutrition, and physical activity in a single area graph (Figure 1b). To increase the ecological validity of interpreting data, we visualized actual data recorded from 1 GDM woman before the study. The glucose data (mmol/L) were recorded using Medtronic's continuous glucose meter Enlite [32], nutrition data (macros in grams) were acquired from a food diary (kept for 3 days) and validated by dietitians, and physical activity data (steps) were recorded with a Garmin Vivosmart 3 activity bracelet. The habitual visualization was provided by showing the most recent values on the first page of the app. A small arrow (Figure 1a) indicating the trend in recent blood glucose levels was added. In addition, to follow customized habits, we designed a view where participants could track their habits (Figure 1c) against goals they had set up.
6. Integrate the app with normal pregnancy and existing health care services.	The app showed the gestational weeks on the entry page (Figure 1a), pregnancy information (Figure 1d), and recommendations by a health care professional (Figure 2a) related to this desirable feature.

[a]GDM: gestational diabetes mellitus.

Figure 1. (a) Main page showing pregnancy weeks and recent self-tracking data, (b) data visualization view (bringing self-tracking data into 1 view), (c) habit formation and tracking tool, and (d) general information about pregnancy, the fetus, and GDM. GDM: gestational diabetes mellitus.

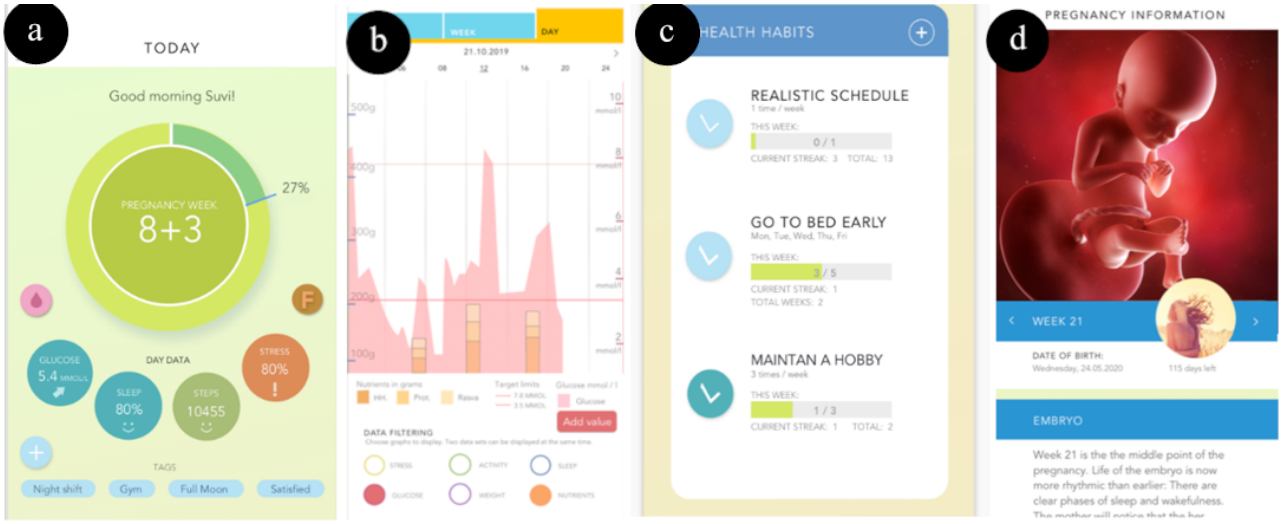
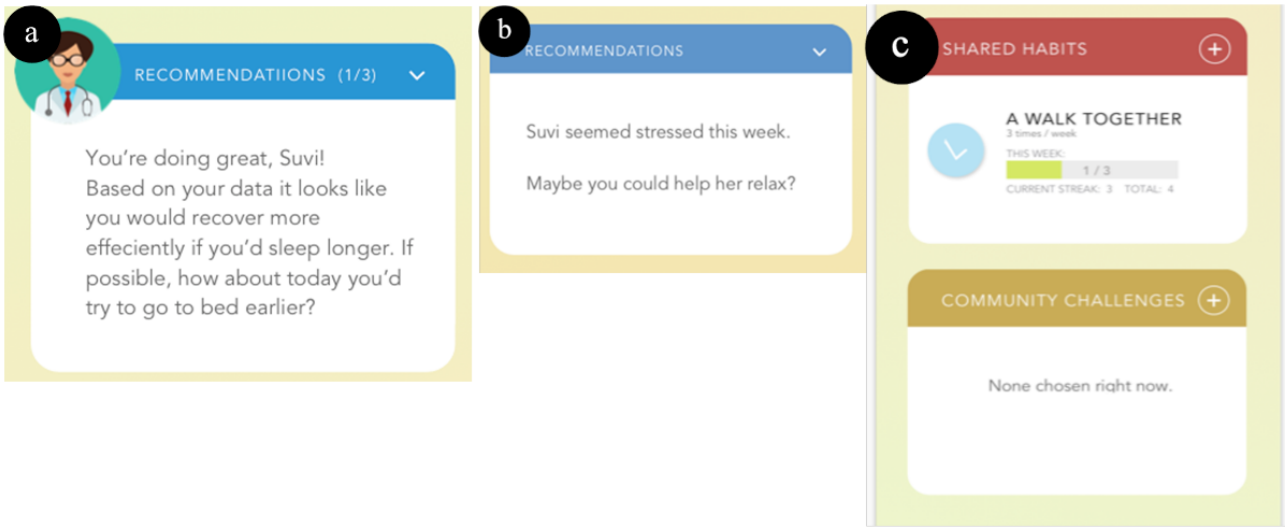


Figure 2. (a) Written reinforcement and feedback, (b) a suggestion for the partner, and (c) shared habits and a community challenge view.



Recruitment and Data Collection

Our goal was to recruit 10 women with GDM to the study from maternity and antenatal clinics in the Helsinki Metropolitan Area (Finland). The clinic nurse asked women with GDM at least at 24 gestational weeks about their interest in participation. If interested, the study nurse contacted the mother with more information about the study and confirmed eligibility. Exclusion criteria were type 1 or type 2 diabetes, use of medication that influences glucose metabolism (eg, oral corticosteroids, metformin, insulin), a GDM diagnosis in previous pregnancies, current substance abuse, severe psychiatric disorder, significant difficulty in cooperating (eg, inadequate Finnish language skills), and significant physical disabilities that prevent the use of a smartphone or moving without aid.

Data were collected using the following procedure. After obtaining informed consent, we collected background information (eg, age, pregnancy weeks, and familiarity with mobile apps) through a questionnaire. Next, the research assistant first briefly presented the idea of the prototype and its limitations (it was possible to move between different screens, but the features were not fully implemented) and then instructed the participant to freely browse the prototype (see the Evaluated GDM Prototype App section) and encouraged her to think aloud [33] and voice any possible questions or comments that might come to mind while doing so. This free exploration of design features was chosen as we wanted to gather experiences of the features rather than examining in detail the usability conducting specific tasks with the app. This is a typical approach in user experience studies [34]. If the participant did not find all the screens available in the prototype, the research assistant showed her the screens she had missed. The research assistant asked questions about the features, and after the participant evaluated each feature, the research assistant asked more general questions about the app (see Textbox 1 for the main interview questions). The research assistant also asked follow-up questions when elaboration of the participant's answer was needed. Regarding desirable features related to social support from the partner (desirable feature 3), we asked participants to evaluate the "femininity" of the app on a scale from 1 (not feminine at all) to 10 (very feminine), as low femininity has been argued to be important for male users [26].

Sessions were conducted in quiet places that were easiest for the participants to travel to and were conducted in their native language. Interviews lasted approximately 1 hour. Interviews were audio-recorded and transcribed. Quotes provided in the results were translated to English verbatim, and filler words, such as "er," were removed during translation.

Textbox 1. Main interview questions.

Main questions about features

- What do you figure out from this feature? What do you think about it?
- What would you like to have in this feature? How would you develop it?
- Would you need help using this feature?

Main questions about the app

- What do you think about the app?
- How do you think the app could help in managing blood glucose?
- How would it feel to use this app in everyday life?
- What do you think about the number of features? Too many or too few? Why?
- What do you think about the look and feel of the app? How feminine do you perceive it to be?

Analysis

Interviews were transcribed, and 2 researchers familiarized themselves with the interviews by reading the transcripts. The analysis was conducted according to the Framework method, which is a recommended approach for multidisciplinary health research [35]. First, 2 researchers read the interviews and coded them using desirable features (see Table 1) as predefined codes, as they were the focus of the analysis. In cases when content outside these codes emerged, we added new codes. The coding was performed using ATLAS.ti [36]. We ended up with 39 codes and grouped them thematically into larger categories. In the Results section, we present the categories in relation to the desirable features and indicate the frequency of each category.

Results

Participants

In total, 10 women with GDM were recruited to the study and interviewed. Table 3 presents their background information. The age of the participants varied from 24 to 40 years (mean 33.6 years, SD 4.4 years), and the gestation weeks varied from 27.1 to 37.0 weeks (mean 33.6, SD 2.7 weeks). Participants were familiar with using various mobile apps, as they agreed with the statement "I am used to using various mobile apps," with a mean score of 4.3 out of 5 (SD 0.9). The mean scores for the statements "I am used to using physical activity sensors (eg, Fitbit, Vivosmart, and Polar)" and "I am familiar with measuring blood glucose" were 2.4 (SD 1.4) and 4.5 (0.5), respectively.

Table 3. Background information about women with GDM[a] who participated in the study. Regarding the statements, participants responded on a Likert scale (1=strongly disagree to 5= strongly agree).

Participant ID	Age (years)	Weeks of gestation, n	Likert score for "I am used to using various mobile apps."	Likert score for "I am used to using physical activity sensors (eg, Fitbit, Vivosmart, and Polar)."	Likert score for "I am familiar with measuring blood glucose."
1	36	35.0	4	3	5
2	32	33.3	4	2	4
3	40	31.2	4	2	4
4	24	33.7	5	2	5
5	31	35.6	4	2	4
6	31	30.3	5	1	4
7	32	36.6	2	1	5
8	36	37.0	5	5	5
9	35	34.8	5	1	4
10	39	28.1	5	5	5

[a]GDM: gestational diabetes mellitus.

Increased Competence to Manage GDM With Automatic Feedback and Interactive Exploration

Most of the participants (9/10, 90%) wished to obtain feedback about their lifestyles. Although all 10 (100%) participants actively followed their blood glucose levels, most of them (8/10, 80%) would have liked to receive more feedback on how to influence blood glucose levels.

> *Well, I would like to follow my general vitality, activity, sleep…like how stressed you are, do you eat well or not, are you able to keep a diary, so that your general wellbeing would be followed, not just the glucose slavishly but how to influence it.* [Participant 8 (P8)]

As the feedback they received was mainly negative, such as whether blood glucose levels are too high, 7 (70%) participants wished to also obtain positive feedback beyond physical activity

> *Well, the Garmin Vivosmart gives you feedback that you have achieved a goal, so a similar thing would be useful, that you would receive some kind of badge that everything has gone well.* [P1]

These results indicate that there is a lot of room for improvement in the amount and quality of the feedback to support self-discovery and further healthy behavior.

Many of the participants (7/10, 70%) wished to obtain clear suggestions on what to eat.

> *I wish it would give precise food suggestions, like what would be a good evening snack if I am having troubles with my fasting glucose in the mornings…so it would follow my diet and give me transformative suggestions so that the glucose values would stay within the limits so that it would be able to tell something like, now eat more protein but nothing else, or something like that.* [P9]

So, in addition to Harrison et al [37], who argued that women with GDM wish to obtain clear suggestions for physical activity, women with GDM wish for suggestions on nutrition as well. Ideally, these suggestions would be personalized according to the user's diet.

> *I would like to have more tailored food suggestions, for example, if you are vegan…you could suggest*

> *replacing the oat milk with something else or I do not know…I miss out at least a half of these nutrition guides, as they are like, "Eat a rye bread sandwich," but I can't because I can't eat rye bread.* [P1]

However, these suggestions considered not only the contents of the food but also the timing when one should eat. One participant implemented this by herself using alarms on her mobile phone.

> *I have had alarms "Remember to eat" on my mobile phone every two to three hours.* [P10]

In addition, 3 (30%) participants said that the suggestions should be provided gently.

> *For most of the participants, those tips must be useful. So if you replace that one with that one…at least I would be ready to receive that kind of tips, but I would not like to get a sermon; that does not help anything.* [P8]

Most of the participants (8/10, 80%) discussed that it would be motivating if they were able to find positive causes and effects of their behavior and blood glucose levels.

> *I think it would be insanely motivating if I were able to see that my glucose would stay within limits because I went for a walk.* [P1]

However, personalized recommendations do not currently exist in GDM apps.

> *Now you can't get recommendations that walk for 45 mins and your blood sugar would go down. That would very valuable.* [P3]

Most of the participants (8/10, 80%) also requested precise and clear suggestions on lifestyle choices that they can control.

> *But could there be a feature in the app that would really say, you should eat in half an hour. Because that would be really good. And if my observation holds that even a small amount of physical activity decreases or balances blood sugar so it would be good if it said that "go for a walk."* [P3]

This feedback should be provided gently and should respect the autonomy of women with GDM.

> *Here is this kind of recommendation that you would recover better by having longer sleep, having this kind of concrete suggestion when you are feeling tired with data that you slept badly would help. And the suggestion is not provided as an order that goes to sleep immediately…so it is given in a pretty pleasant form.* [P6]

The self-discovery process of finding how lifestyle influences blood glucose levels is tedious [8]. The app could facilitate this process by providing feedback and suggestions on what the person could try.

> *It would be good if you could get this kind of conclusion on what to try next, because now when you are figuring it out yourself, it is really like hit-and-miss what to try next…was it the sleep, or two pieces of bread or egg. So if it [the app] could help in this, it would be really great.* [P9]

Although our approach to providing recommendations was appreciated by 8 (80%) participants, providing feedback about sleeping, in particular, was not found to be useful.

> *I have only three and half weeks until the due date. I have no control oversleeping, for example.* [P1]

Thus, the feedback should be directed toward lifestyle changes that can be influenced by women with GDM.

To conclude, feedback (also positive) about lifestyle was highly appreciated, but it should be clear, feasible, and gentle. Feedback should also be directed toward lifestyle changes that women with GDM can influence, especially toward diet and physical activity by encouraging women to try different options for changing their diet and increasing physical activity. Suggestions for sleeping habits were not perceived as useful.

Increase Autonomy by Enabling Personalization

Most of the participants' (8/10, 80%) responses supported the importance of personalization.

> *Otherwise this looks good, but it would be good if you could modify this more based on one's own needs so that if someone has problems with sugar levels and someone with physical activity…so everyone would be not be cut from the same cloth, and it could be slightly adjusted based on what one is able and wants to do.* [P7]

Personalization of habits was found to be motivating by over half (6/10, 60%) of the participants.

> *It would be motivating if you were able to add your own daily tasks, that I prefer more.* [P3]

Moreover, further evidence for the need for personalization of habits was revealed, as 3 (30%) participants found the proposed habits annoying.

> *I have to say about that "Maintain a hobby." Being at this last stage [of pregnancy] is very laborious, and you know what kind of lifestyle you should have, but I don't have enough strength really…I can't have a hobby, so I am really annoyed by this kind of thing.* [P3]

Beyond our implementation, 7 (70%) participants discussed the personalization of goals for lifestyle choices. For example, the goal for physical activity should be customized to be achievable.

> *If I would realize at first that I am able to walk only 2000 steps a day, the 10 000 steps are pretty far. So if you could adjust the goal to be like 5000.* [P3]

Personalization was discussed for nutrition, as participants indicated that customized nutrition suggestions are important (see the Increase Competence to Manage GDM With Automatic Feedback and Interactive Exploration section).

The possibility to add a name, profile picture, and expected date of birth was not extensively discussed, although 3 (30%) participants preferred personal greetings as they made them feel nice.

> *Somehow this "Good morning, Suvi" feels nice, it feels personal…this is like a modern diary...so I would personally like to use this for sure.* [P4]

To conclude, participant responses reflected the relationship between personalization and autonomy. The ability of women with GDM to follow health recommendations, especially in terms of physical activity and sleeping, varies significantly between individuals and stages of pregnancy. Thus, these recommendations, user profiles, habits, and goals should be highly customizable. As these personalized recommendations are obtained, at least different options (eg, how to increase physical activity and what to eat) should be provided.

Provide Social Support, Especially From the Partner

Incorporating this desirable feature raised divided opinions from participants. We observed that participants without previous children liked the functionality.

> *That walking together is very nice…maybe that is probably my favorite.* [P4]

However, participants who had given birth before were skeptical about this feature.

> *I think that in our case the partner's/caregiver's part in the app wouldn't be used. Maybe for people who are pregnant for the first time, for whom all this is very new…so, wouldn't be our thing, but [a] nice feature in general.* [P8]

Although the majority of the participants (6/10, 60%) preferred that their partners had a version of the app, the support needed from the partner varied. Some participants (4/10, 40%) discussed the role of the partner as providing support, such as a reminder to go sleep earlier or to provide encouragement in eating.

> *So it [the app] would notify my partner like [participant's name] has slept badly, so ask her to go to bed earlier…that kind of additional feature would be good.* [P4]

> *It is good that the partner can see the weeks of gestation, and when thinking of gestational diabetes. It could remind my partner to encourage me in eating or something like that.* [P5]

As such, some of the reminders could originate from the partner, which could be more effective than providing reminders only for women with GDM. However, further studies are needed to investigate how women with GDM would perceive this.

> *Well, this [gestational diabetes] is also a bit of partner's thing, or at least he gets influenced when I am saying aloud that my blood sugar is there so what should I do…my partner has also tried being on this same diet. I don't know if it becomes oppressive if he monitors me; I do not know. But I think it would be nice if he could suggest today this and that from there [the app].* [P9]

In addition, 3 (30%) participants discussed that their partners would be willing to help if the reasons were clear.

> *At this stage, when the state of health is not the best, he [partner] starts to understand the realities, that something needs to be done and learned. So, I am sure he would receive the instructions and would be able to act accordingly if the reasons were justified well.* [P10]

However, 1 (10%) participant who did not find the partner's view in the app useful stated that the app could be used to prove the changed conditions due to pregnancy.

> *You could use this as a weapon at home: "Look how little I have slept, so I need to sleep longer in the morning."* [P8]

We also evaluated the femininity of the prototype app, as described in the Recruitment and Data Collection section. Although the participants rated the app's femininity to be rather high (median 7), the ratings on femininity varied considerably (minimum=3, maximum=8). The ratings were more influenced by the topics related to pregnancy

> *Well, there is a picture of a baby, so it is probably the only feminine thing here.* [P5]

rather than the feel of the user interface:

> *There are nothing like stereotypical things, like flowers, hearts, and butterflies, which I have seen in period apps.* [P6]

The neutral and somewhat clinical look was preferred by the participants.

> *The feeling of medical science, makes it feel credible...if it was fluffy pink, I would not use it myself.* [P1]

The use of neutral elements and colors would support the usage of the app by both women with GDM and partners.

The community challenge raised fewer opinions. Of the 10 participants, 2 (20%) discussed belonging to a peer support group on social media and 1 (10%) participant found that this support could be improved by having things that women with GDM could do together.

> *I am part of a peer support group in Facebook, I have not been very active there, but people think about these things [issues about GDM] a lot and some things are left on their own, so it could be interesting if something like this is supported, at least like in this way virtually.* [P9]

However, most participants (8/10, 80%) did not comment on this feature.

To conclude, incorporating social support into the app divided opinions, and participants suggested different supportive roles for partners. The partner's role in the app should be optional, and the partner should have their own version of the app. This partner's version could include encouragement messages and reminders for a healthier lifestyle, which the partner can communicate with the woman with GDM. Some participants wished to share information about GDM with their partners so that the partners could support and help them in the management of GDM. The more advanced features, such as sharing data and suggestions for common activities, should be customizable.

Support a Normal Pregnancy and Debunk Myths About GDM

Most of the participants (8/10, 80%) preferred having pregnancy information in the GDM app.

> *It is always nice to read what is the situation of the baby.* [P3]
>
> *I think that one can never have too much information about these things.* [P4]
>
> *I have this pregnancy app on my phone...I have looked at it almost twice a day from the beginning…this is something that I am interested in…how much it weighs and what parts have developed this time…so super big plus having this.* [P8]

This could be developed further by providing information not only about pregnancy but also about nutrition and physical activity.

> *Here is information about physical activity and pregnancy, [a] nutrition guide, information about gestational information…this is interesting, so it is more information than what I have seen from apps that I am using.* [P6]

To conclude, the participants valued having pregnancy and GDM information in the app. In addition, they wanted more information about nutrition and physical activity than what is available in current pregnancy apps. Optimally, this information should be personalized together with lifestyle suggestions (see the Increase Autonomy by Enabling Personalization section).

Support Dual Processing as Pregnancy Is Life Changing

In general, almost all the participants (9/10, 90%) found our approach useful for supporting comprehensive self-tracking and to visualize these data in a single graph.

> *This seems somehow useful, as it is not only my own inference between two different graphs, so if I move then I would see how the blood sugar looks like, so having these graphs together seems like a good idea.* [P6]
>
> *Things that I like to follow: blood sugar, sleep, activity, those all are nicely and clearly shown together.* [P2]

Participants also embraced the possibility to select what factors to view in the visualization.

> *I like that you can choose in the graph which ones are visible, then you can compare blood sugar to activity or sleep or something similar.* [P6]

However, some participants (4/10, 40%) discussed that the possibility to select the factor did not simplify the view enough, as they thought that there was too much information and the view was too complex (eg, blood glucose information should be more distinct).

> *I think this [blood glucose] should be emphasized, visually it looks more like a background…so this glucose measurement is the most important thing, and these others are additional information.* [P1]

In addition, 1 (10%) participant requested highlighting blood glucose and having other data as an additional option.

> *In the graphs, there was too much information at a first glance. So, you could simplify this…well you can select to view the data that the user is interested in…so viewing all the graphs at the same time could be one possibility but maybe not the default option.* [P5]

Despite these concerns, the participants indicated that they would have benefited from such visualization at the beginning when they received the GDM diagnosis.

Concerning habitual visualization, 9 (90%) participants liked to have the glanceable data on the first page of the app. The importance of providing an arrow with blood glucose for indicating the trend (see Figure 1a) was also identified here.

> *I think those are great, like the glucose, there is an arrow, so it gives more information for what to compare.* [P5]

Some participants (4/10, 40%) indicated that they should pay attention to the app only when something is going wrong and thus support more reactive decision-making.

> *I think that the most important thing in the app would be like "Mom, relax. It will alarm when you need to do something about it."* [P1]

Reminders were also discussed to help create measurement habits.

> *When I am having a post-meal measurement, I always add it as an alarm to my mobile phone. So, of course, if that would be possible using the app, I would not have to do it separately.* [P3]
>
> *It could remind me when I need to measure more than the normal blood sugar in the morning.* [P8]

The mobile phone was mentioned as being useful for tracking habits, as people follow their mobile phones anyway.

> *For example, the goals, like habits and so on, what you can put there…that feels something that would be useful for me. So when you add them and the app would remind me, it would be something that I would follow, because I follow the mobile phone quite often. If you had a separate diary or something like a paper or calendar, then you would not most likely notice them.* [P6]

As such, in addition to customizing the habit (see the Increase Autonomy by Enabling Personalization section), the reminders of these habits could also be customized. A participant discussed that once healthier habits have become part of everyday life, assistance from the app would not be needed anymore.

> *You would need to follow the app every day pretty slavishly…so I think one of the goals of this app would be to learn to be without it…the less you would need it, the better. Now it is quite engaging…so there*

> *should be a balance…the app is more like help but not something that flogs.* [P7]

To conclude, both approaches (habitual and reflective) to support decision-making should be incorporated in the design. The dual processing responds to the various needs of women with GDM, as some were interested in investigating the self-tracking data in depth, whereas others preferred the glanceable information about current values. It is recommended to use simple visualizations (eg, using the traffic light metaphor or arrows) to show real-time feedback and line charts combining blood glucose levels and lifestyle for providing detailed information about the causes and effects. Participants had different opinions on how much they are willing to pay attention to data and how much should be provided in the form of notifications and alarms.

Integrate the App With Normal Pregnancy and Existing Health Care Services

In addition to supporting desirable features, providing information that is common for pregnant women in general and women with GDM supported the feeling of normal pregnancy. Having the gestational weeks on the first page was clear for the participants.

> *Pregnancy weeks are easy to understand.* [P1]

> *The overall appearance of the app looks clear as there are pregnancy weeks on the first page.* [P7]

Providing pregnancy information in addition to self-tracking data was appreciated by the participants (8/10, 80%).

> *It's nice that you can go from the first page to pregnancy information, so everything is not just hard data, but also information similar to pregnancy applications.* [P10]

Almost all the participants (9/10, 90%) suggested that the app could be part of health care, especially if sufficient instruction is provided.

> *I think this should be some kind of maternity clinic app right away, and good instruction should be provided.* [P1]

> *A pretty good introduction should be given at the start, and then there could be a possibility to quit using it [the app] if it doesn't feel one's thing. If you are new and suddenly you are given this kind of app, it can be that it doesn't suit everyone.* [P7]

Participants discussed the importance of having contact with a diabetes nurse so that they can share the data with them and discuss the data provided by the glucose sensor.

> *I think that a chat and feedback functionalities would be useful, so it would send an alarm if you get a lot of hypers, and that information should be automatically transmitted to the diabetes nurse…I think the current way is very old school…now we send some glucose values by email, which can basically be anything as they can vary from time to time.* [P1]

Many of the participants (7/10, 70%) wanted to receive a message from the app to contact a clinic if something unusual occurs.

> *There should be contact details for maternity clinics or something like that. So, when you face some challenges, for example, if you have had three hypers this week, you would get a notification to contact the clinic.* [P4]

More than half of the participants (6/10, 60%) also mentioned that the app should be integrated with visits to a clinic and the app could support collecting more data and provide them for health care personnel.

> *You could utilize the app in maternity clinics so that there weren't so many different units…that you could combine this to maternity clinics, so this could be a good package for them. Then you could also add blood pressure or weight or something to the app you once have visited the clinic.* [P7]

To conclude, participants liked to have gestational weeks on the first page and pregnancy information to support the feeling of normal pregnancy. They also felt that the app could be part of their routine care and provided different suggestions on how the app could be integrated into health care services, including good instructions, use of self-tracking data for visits to maternity clinics, and enabling connection to the maternity clinic when blood glucose levels have frequently been exceeded.

Other Emerged Themes

Most of the comments from women with GDM were consistent with literature-based desirable features [14], indicating that desirable features cover the majority of themes discussed in the interviews. However, as we analyzed emerging themes beyond the literature-based desirable features, we want to note the following important aspects.

Basic Functionalities That Are Simple to Learn but Have Sufficient Depth to Maintain Interest

The temporal dimension (ie, how the use of the app would vary as a function of time) was not extensively discussed in the desirable features. However, this issue was raised in the interviews by 4 (40%) participants.

> *The more I would use this app, the more I would benefit from it…the first week I would be filling information, but then I would start setting habits and other things and would gain more benefits.* [P8]

This emphasizes the need for good instructions when using the app, which was discussed in the Integrate the App with Normal Pregnancy and Existing Healthcare Services section. After the initial stage, the app should have sufficient depth for maintaining interest.

> *There should be more features rather than less, so it is versatile. And it [the app] is a sort of combination of things that I would use anyway, so having a lot of features doesn't bother me…it maintains interest if there are many different things.* [P4]

As such, provided that the usability is not compromised due to multiple features, having them would be beneficial. This seemed to be the case for our app.

> *This [app] is pretty clear and simple…and after using it a couple of minutes this feels easy to learn and grasp…when compared to other apps I'm using and have used.* [P10]

Tolerance for Slipping and Usage Breaks

Half of the participants (5/10, 50%) discussed that the app should be started on the mother's own terms (see also the Increase Competence to Manage GDM With Automatic Feedback and Interactive Exploration section regarding starting self-tracking). This includes tolerance for slipping, which happens to women with GDM as well.

> *I would consider this app interesting as I slip sometimes…but it [the app] should not increase anxiety if these slips are recorded there…No one is perfect, and there could be some things that you at least won't be blamed for if you slip…If you need to show this to someone, you might neglect some markings or not use it if you feel ashamed that you haven't been that good. So, having that kind of app would be the best option.* [P7]

This emphasizes that the app should focus on the positive aspects of how person is managing GDM, for example, through positive feedback (see the Increase Competence to Manage GDM With Automatic Feedback and Interactive Exploration section). In addition, 3 (30%) participants discussed that their attitude toward the app might vary time to time.

> *For most of the time, the app would be like a good thing, but I can imagine that there would be times that the whole app would start annoying me.* [P2]

This indicates that the app should be tolerant of usage breaks and different usage frequencies.

Discussion

Principal Findings

In this study, we examined how to incorporate literature-based desirable features [14] into mobile apps designed for management of GDM. Most of the feedback on example implementations of the desirable features was positive and constructive. We identified that self-tracking data in GDM apps should be extended with written feedback, habits and goals should be highly customizable to be useful, providing social support through the app should be possible for the partner, and health care professionals should be notified through the app if something unusual occurs. Relating to automatic feedback and self-tracking data exploration (desirable feature 1), the feedback through the self-tracking data was of value, but the interviews revealed that the feedback should be extended with written feedback. The interviews also indicated that the guidance should be clear, feasible, and gentle. In general, these comments are consistent with findings by Harrison et al [37], which indicated that women with GDM want clear suggestions on what physical activity to engage in. However, this should be provided gently and should consider personal physical capabilities and preferences for diet, respecting the women's autonomy. Many women encounter difficulties in trying to adapt ethnic diets to the dietary changes that GDM demands [27,38,39]. Draffin et al [27] considered giving culturally appropriate advice as crucial to prevent women from feeling alienated, as it can be difficult for women to follow recommendations when they feel that such recommendations are not applicable to their lives. Providing automatic guidance in the form of recommendations or other written feedback is largely missing in existing GDM apps [14]. Thus, future work is needed on how to provide influential feedback for women with GDM with automatic approaches to achieve the effectiveness of human feedback [14].

Regarding the personalization possibilities in the app (desirable feature 2), our results support findings by Goetz et al [40], who reported that personalized welcome messages and the possibility to add their pictures were appreciated by a third of pregnant women. However, participants stressed the importance of personalizing daily habits and goals, which has not been possible in existing GDM apps [41]. As such, there is a lot of room to enhance the personalization features of GDM apps. Personalization should not be limited to lifestyle choices but should also include personal concerns related to pregnancy [38]. How to personalize apps for GDM for women from different ethnic and cultural backgrounds in managing their diabetes requires further studies.

Social support (desirable feature 3) has also been absent, as none of the existing GDM apps provides social support from the partner [14]. The responses from participants indicated that their partners could have different supportive roles through the app. More investigations are required to determine how this could be implemented such that women with GDM do not feel controlled but are supported in making healthier lifestyle decisions. Showing simple notifications to the partner, without accessing the woman's self-tracking data, would be a good starting point as the amount of data disclosed to the partner would be minimized. The user could then decide whether they would be willing to share more data with their partner. In addition, the partners' opinions about the intended functionalities for them should be studied more. These investigations have been conducted for pregnancy apps designed for men [42] but are lacking in GDM apps [14]. Beyond partners, providing social support through a community challenge did not raise many comments from the partners. This is consistent with the findings by Peyton et al [26], who showed that more research is needed on how to incorporate social media activities in pregnancy apps.

Regarding debunking myths about pregnancy (desirable feature 4), the responses were consistent with the literature that women with GDM want to have GDM and pregnancy information in the same place [17]. However, participants requested more in-depth information about nutrition and physical activity than the current apps provide, akin to a previous study [17]. As such, GDM apps should have a comprehensive information section about pregnancy, GDM, and related topics, such as nutrition (eg, recipes) and physical activity.

Regarding the depth of information, the participants' willingness to interpret the self-tracking data significantly varied. Some wanted to discover cause and effect, whereas others just wanted to know "if something is wrong." This indicates that support for dual processing (desirable feature 5) is useful. The

dual-process approach is supported fairly well in current GDM apps in terms of providing data at varying levels of detail [14]. Habitual processing is implemented with glanceable visualization of recent blood glucose levels, whereas more detailed graphical visualization is provided to support reflective processing. Habitual or reactive processing is further supported with reminders and alarms. However, the possibility to create and track daily habits is missing from current GDM apps [14]. Participants would especially appreciate habit tracking if they were able to personalize their habits (see desirable feature 3). Moreover, linking the new or changeable habit to an existing habit can be expected to be a more effective approach than time- and phone-based approaches in the long term. How this could be accomplished remains an open challenge, but as the techniques for activity recognition continue to develop [43], detecting the existing habits may soon become possible, thus enabling linkage of new and changeable habits with existing habits, such as measuring blood glucose levels, which is a frequent task for women with GDM.

Finally, regarding desirable feature 6, integration to a normal pregnancy was appreciated as participants discussed having gestational weeks on the first page as a good and familiar feature from pregnancy apps. Participants also proposed ways for how the app could be integrated into health care services in terms of monitoring between maternity clinic visits. These findings are in line with the review on technological support for diabetes management, which emphasizes the importance of 2-way communication between people with diabetes and health professionals [29] and that the self-tracking with wearable sensors can increase the completeness of the self-tracking data presented to health care professionals [44]. However, it is unclear how health care professionals would perceive self-tracking data beyond blood glucose levels. A big issue on acceptability is their attitude toward the data provided by lifestyle and well-being devices [44,45]. These acceptability investigations of well-being data among health care professionals are missing in the context of GDM. This integration has been partly implemented in current GDM apps [46,47], and the absence of this integration is considered a large shortcoming [17].

In addition, we found 2 important themes from the interviews that were not clearly related to desirable features. First, as GDM is perceived as “a stun” [48], the participants indicated that the app should be simple to take into use. This is particularly important at the beginning, when some effort is required for setting up the app. There should be also deeper features to maintain interest for women with GDM at a later stage of pregnancy. We believe that it is possible to design an app that is easy to use but has many features. This is in contrast to the findings by Hood et al [49], who observed that the number of features decreases the usability of diabetes apps in general. However, keeping the basic functionalities simple and having sufficient depth from additional features seem to be a good approach. Second, as women with GDM may have feelings of guilt, the app should have a tolerance and supporting approach for slipping and unfavorable behavior. Skar et al [17] discuss that slipping might lead to undesirable “cheating.” This also implies that returning to the app after a while should be easy and that lack of use should not be punished but rather supported by friendly messages to guide women with GDM to start using the app again.

Limitations of the Study

We are aware that the design space for implementing each desirable feature is large, and thus, multiple different implementations can be derived from it. At this stage, we used these example implementations of desirable features as starting points to explore and evoke further ideas from participants. As such, our aim was not to validate these implementations of the desirable features but to gain constructive feedback to incorporate the desirable features into a future GDM app.

We acknowledge that the number of participants could have been higher. However, we believe that this was sufficient as the data seemed to be saturated after 8 interviews. Moreover, the same number of participants have been used in qualitative studies on experiences of GDM (eg, [48]).

The participants were familiar with using mobile apps, which reflects the wide adoption of mobile apps among pregnant women in developed countries [50]. However, more work is needed to assess how acceptable and feasible the app-based interventions are for women with GDM in large-scale use [18].

Directions for Future Work

Most of the studies on GDM apps have investigated the effects of these apps on medical outcomes or user experience. However, less work has focused on constructive approaches for investigating the design features of GDM apps. This study aimed to fill the gap by designing and providing examples of features that have justifications in the literature [14] and by providing the perspectives of women with GDM. In the future, these features should be implemented and integrated into a functional GDM app to investigate their impact on GDM self-management. More research is required to understand the efficacy of each exploratory desirable feature provided here. To do this, we will incorporate at least part of these features in a fully functional mobile app for iOS and Android and conduct a randomized controlled trial to evaluate its effect on maternal and neonatal outcomes in the future. In addition, the feasibility of implementing each desirable feature is largely unknown. However, the implementation of providing reliable information can be considered to be much easier than the implementation of habit formation and tracking, for example.

An interesting part of future work is to study interactive learning environments with women with GDM that allow creating causalities between health behaviors (diet, physical activity, and sleep) and blood glucose levels. Especially, this would support desirable features 1 and 4. A survey on the features of diabetes apps by Adu et al [51] revealed that the teaching of diabetes self-management skills is underrepresented in diabetes apps [52], although recent guidelines emphasize the need for patient education [53]. Qualitative studies [21,27] indicate that the self-discovery process of GDM is challenging and demanding, which takes a considerable amount of time. Nevertheless, the self-discovery process could be expedited and facilitated. For example, this could be achieved with interactive visualizations of how lifestyle impacts blood glucose levels and

what kind of potential impact this can have on the child. Bień et al [54] noted that increased knowledge of diabetes during pregnancy increases perceived general health. Carolan-Olah et al [55] suggested that knowledge of GDM could be improved, particularly for women from multiethnic and low socioeconomic backgrounds. Although an app is a common way to provide an eHealth intervention, the interactive learning environment does not necessarily need to be restricted to apps. An interactive learning environment could be leveraged to other technologies for diabetes management, as explored by Rollo et al [56]. These include web-based programs, video games, and even emerging technologies, such as virtual reality and augmented reality. For example, the use of virtual reality can provide an immersive environment to rehearse healthier choices [56]. However, it is still unclear what benefits these technologies could bring specifically to GDM management over mobile health (mHealth) learning environments.

Conclusion

This study was intended as a constructive step in the design and development of a mobile app to support behavior change required in the self-management of GDM. We designed and evaluated how to incorporate desirable features into GDM apps. The results confirmed the importance of these features to support self-management through GDM apps, and we gained constructive information about the contents and functionalities of each feature. We expect that the implementation of at least some of these features will increase the efficacy of GDM self-management apps. We argue that the results are useful for various stakeholders that target effective mHealth interventions for GDM.

Acknowledgments

We wish to thank all the women with GDM in Finland taking part in the research and the study nurses Riikka Luumi, Milla Tuhkanen, Jaana Palukka, and Sanna Lampi. This work was funded by the Business Finland eMOM gestational diabetes mellitus (GDM) project.

Conflicts of Interest

None declared.

References

1. Wang H, Li N, Chivese T, Werfalli M, Sun H, Yuen L, IDF Diabetes Atlas Committee Hyperglycaemia in Pregnancy Special Interest Group. IDF Diabetes Atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by International Association of Diabetes in Pregnancy study group's criteria. Diabetes Res Clin Pract 2022 Jan;183:109050. [doi: 10.1016/j.diabres.2021.109050] [Medline: 34883186]
2. Burlina S, Dalfrà MG, Lapolla A. Long-term cardio-metabolic effects after gestational diabetes: a review. J Matern Fetal Neonatal Med 2021 Mar 28:1-8. [doi: 10.1080/14767058.2021.1903863] [Medline: 33779467]
3. Ornoy A, Becker M, Weinstein-Fudim L, Ergaz Z. Diabetes during pregnancy: a maternal disease complicating the course of pregnancy with long-term deleterious effects on the offspring. a clinical review. Int J Mol Sci 2021 Mar 15;22(6):2965 [FREE Full text] [doi: 10.3390/ijms22062965] [Medline: 33803995]
4. Farrar D, Simmonds M, Bryant M, Sheldon TA, Tuffnell D, Golder S, et al. Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. BMJ 2016 Sep 13;354:i4694 [FREE Full text] [doi: 10.1136/bmj.i4694] [Medline: 27624087]
5. Bellamy L, Casas J, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009 May 23;373(9677):1773-1779. [doi: 10.1016/S0140-6736(09)60731-5] [Medline: 19465232]
6. Carolan M, Gill GK, Steele C. Women's experiences of factors that facilitate or inhibit gestational diabetes self-management. BMC Pregnancy Childbirth 2012 Sep 18;12(1):99 [FREE Full text] [doi: 10.1186/1471-2393-12-99] [Medline: 22988897]
7. Brown J, Alwan NA, West J, Brown S, McKinlay C, Farrar D, et al. Lifestyle interventions for the treatment of women with gestational diabetes. Cochrane Database Syst Rev 2017 May 04;5:CD011970 [FREE Full text] [doi: 10.1002/14651858.CD011970.pub2] [Medline: 28472859]
8. Craig L, Sims R, Glasziou P, Thomas R. Women's experiences of a diagnosis of gestational diabetes mellitus: a systematic review. BMC Pregnancy Childbirth 2020 Feb 07;20(1):76 [FREE Full text] [doi: 10.1186/s12884-020-2745-1] [Medline: 32028931]
9. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: a global overview. Diabetes Metab Res Rev 2018 May;34(4):e2990. [doi: 10.1002/dmrr.2990] [Medline: 29431916]
10. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 06;20(1):198 [FREE Full text] [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]
11. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]

12. Rigla M, Martínez-Sarriegui I, García-Sáez G, Pons B, Hernando ME. Gestational diabetes management using smart mobile telemedicine. J Diabetes Sci Technol 2018 Mar;12(2):260-264 [FREE Full text] [doi: 10.1177/1932296817704442] [Medline: 28420257]
13. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
14. Kytö M, Strömberg L, Tuomonen H, Ruonala A, Koivusalo S, Jacucci G. Behavior change apps for gestational diabetes management: exploring desirable features. Int J Hum–Comput Interact 2022 Nov 07;38(12):1095-1112. [doi: 10.1080/10447318.2021.1987678]
15. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
16. Garnweidner-Holme LM, Borgen I, Garitano I, Noll J, Lukasse M. Designing and developing a mobile smartphone application for women with gestational diabetes mellitus followed-up at diabetes outpatient clinics in Norway. Healthcare 2015 May 21;3(2):310-323 [FREE Full text] [doi: 10.3390/healthcare3020310] [Medline: 27417764]
17. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
18. Fantinelli S, Marchetti D, Verrocchio MC, Franzago M, Fulcheri M, Vitacolonna E. Assessment of psychological dimensions in telemedicine care for gestational diabetes mellitus: a systematic review of qualitative and quantitative studies. Front Psychol 2019 Feb 5;10:153 [FREE Full text] [doi: 10.3389/fpsyg.2019.00153] [Medline: 30804842]
19. Buchenau F, Suri JF. Experience prototyping. 2000 Presented at: DIS '00: Proceedings of the 3rd Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques; August 17-19, 2000; New York, NY p. 424-433. [doi: 10.1145/347642.347802]
20. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key Components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
21. Carolan M. Women's experiences of gestational diabetes self-management: a qualitative study. Midwifery 2013 Jun;29(6):637-645. [doi: 10.1016/j.midw.2012.05.013] [Medline: 22877761]
22. Gao Y, Li X, Lin YH, Liu X, Pang L. Nuwa: enhancing the pregnancy experience for expectant parent. 2014 Presented at: CHI EA '14: CHI '14 Extended Abstracts on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Ontario, Canada p. 257-262. [doi: 10.1145/2559206.2580928]
23. Prabhakar A, Reyes LG, Effron A, Kleinschmidt VM, Driscoll M, Peters C, et al. "Let me know if you need anything": support realities of new mothers. 2017 Presented at: PervasiveHealth ’17: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 2017; Barcelona, Spain p. 31-40. [doi: 10.1145/3154862.3154863]
24. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
25. Symons Downs D, Ulbrecht JS. Understanding exercise beliefs and behaviors in women with gestational diabetes mellitus. Diabetes Care 2006 Feb;29(2):236-240. [doi: 10.2337/diacare.29.02.06.dc05-1262] [Medline: 16443866]
26. Peyton T, Poole E, Reddy M, Kraschnewski J, Chuang C. "Every pregnancy is different": designing mHealth interventions for the pregnancy ecology. 2014 Presented at: DIS '14: 2014 Conference on Designing Interactive Systems; June 14-18, 2014; Vancouver, BC, Canada p. 577-586. [doi: 10.1145/2598510.2598572]
27. Draffin C, Alderdice F, McCance D, Maresh M, Harper R, McSorley O, et al. Exploring the needs, concerns and knowledge of women diagnosed with gestational diabetes: a qualitative study. Midwifery 2016 Sep;40:141-147. [doi: 10.1016/j.midw.2016.06.019] [Medline: 27553869]
28. Katz DS, Price BA, Holland S, Dalton NS. Designing for diabetes decision support systems with fluid contextual reasoning. 2018 Presented at: CHI '18: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Canada p. 1-12. [doi: 10.1145/3173574.3174199]
29. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep 31;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
30. Muuraiskangas S, Mattila E, Kyttälä P, Koreasalo M, Lappalainen R. User experiences of a mobile mental well-being intervention among pregnant women. In: Serino S, Matic A, Giakoumis D, Lopez G, Cipresso P, editors. Pervasive Computing Paradigms for Mental Health. MindCare 2015. Communications in Computer and Information Science, vol 604. Cham: Springer; 2016:149.
31. inVISION. Online Whiteboard Meets Productivity Platform: A Real-Time Workspace for Teams to Collaborate Better, Align Faster, and Move Forward Together. URL: https://www.invisionapp.com, [accessed 2022-10-03]

32. Medtronic. Getting Started with Continuous Glucose Monitoring. Guardian Connect with Guardian Sensor 3. URL: https://www.medtronic-diabetes.com.au/sites/default/files/guardian_connect_with_gs3_getting_started_guide_final_dec_2018.pdf [accessed 2022-10-03]
33. van Someren MW, Barnard YF, Sandberg J. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. Cambridge, MA: Academic Press; 1994.
34. Vermeeren APOS, Law ELC, Roto V, Obrist M, Hoonhout J, Väänänen-Vainio-Mattila K. User experience evaluation methods: Current state and development needs. 2010 Presented at: NordiCHI 2010: Extending Boundaries - 6th Nordic Conference on Human-Computer Interaction; 2010; Reykjavik, Icelandz p. 521-530. [doi: 10.1145/1868914.1868973]
35. 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 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
36. ATLAS.ti. Turn Your Data into Qualitative Insights, Faster and Easier. URL: https://atlasti.com [accessed 2022-10-03]
37. Harrison AL, Taylor NF, Frawley HC, Shields N. Women with gestational diabetes mellitus want clear and practical messages from credible sources about physical activity during pregnancy: a qualitative study. J Physiother 2019 Jan;65(1):37-42 [FREE Full text] [doi: 10.1016/j.jphys.2018.11.007] [Medline: 30573442]
38. Bandyopadhyay M, Small R, Davey MA, Oats JJN, Forster DA, Aylward A. Lived experience of gestational diabetes mellitus among immigrant South Asian women in Australia. Aust N Z J Obstet Gynaecol 2011 Aug;51(4):360-364. [doi: 10.1111/j.1479-828X.2011.01322.x] [Medline: 21806582]
39. Sajjad UU, Shahid S. Baby+: a mobile application to support pregnant women in Pakistan. 2016 Presented at: MobileHCI '16: 18th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 6-9, 2016; Florence. Italy p. 667-674. [doi: 10.1145/2957265.2961856]
40. Goetz M, Müller M, Matthies LM, Hansen J, Doster A, Szabo A, et al. Perceptions of patient engagement applications during pregnancy: a qualitative assessment of the patient's perspective. JMIR Mhealth Uhealth 2017 May 26;5(5):e73 [FREE Full text] [doi: 10.2196/mhealth.7040] [Medline: 28550005]
41. Chen Q, Carbone ET. Functionality, implementation, impact, and the role of health literacy in mobile phone apps for gestational diabetes: scoping review. JMIR Diabetes 2017 Oct 04;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]
42. Thomas GM, Lupton D, Pedersen S. ‘The appy for a happy pappy’: expectant fatherhood and pregnancy apps. J Gender Stud 2017 Mar 16;27(7):759-770. [doi: 10.1080/09589236.2017.1301813]
43. Abdallah ZS, Gaber MM, Srinivasan B, Krishnaswamy S. Activity recognition with evolving data streams. ACM Comput Surv 2018 Sep 06;51(4):1-36. [doi: 10.1145/3158645]
44. West P, Van Kleek M, Giordano R, Weal MJ, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. 2018 Presented at: CHI '18: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Canada p. 1-13. [doi: 10.1145/3173574.3174058]
45. West P, Giordano R, Van Kleek M, Shadbolt N. 2016 Presented at: CHI '16: 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 3066-30678. [doi: 10.1145/2858036.2858445]
46. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1105-1114 [FREE Full text] [doi: 10.1177/1932296814542271] [Medline: 25004915]
47. Peleg M, Shahar Y, Quaglini S, Fux A, García-Sáez G, Goldstein A, et al. MobiGuide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestational diabetes domains. User Model User-Adap Inter 2017 Mar 11;27(2):159-213. [doi: 10.1007/s11257-017-9190-5]
48. Persson A, Winkvist A, Mogren I. 'From stun to gradual balance': women's experiences of living with gestational diabetes mellitus. Scand J Caring Sci 2010 Sep;24(3):454-462. [doi: 10.1111/j.1471-6712.2009.00735.x] [Medline: 19878489]
49. 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] [Medline: 27412774]
50. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
51. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the development of mobile phone apps to support diabetes self-management: systematic review. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]
52. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011 Sep 22;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
53. American Diabetes Association. Standards of medical care in diabetes-2016 abridged for primary care providers. Clin Diabetes 2016 Jan 15;34(1):3-21 [FREE Full text] [doi: 10.2337/cd16-0067] [Medline: 28144042]

54. Bień A, Rzońca E, Kańczugowska A, Iwanowicz-Palus G. Factors affecting the quality of life and the illness acceptance of pregnant women with diabetes. Int J Environ Res Public Health 2015 Dec 22;13(1):ijerph13010068 [FREE Full text] [doi: 10.3390/ijerph13010068] [Medline: 26703697]
55. Carolan-Olah M, Steele C, Krenzin G. Development and initial testing of a GDM information website for multi-ethnic women with GDM. BMC Pregnancy Childbirth 2015 Jul 05;15:145 [FREE Full text] [doi: 10.1186/s12884-015-0578-0] [Medline: 26142482]
56. Rollo ME, Aguiar EJ, Williams RL, Wynne K, Kriss M, Callister R, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. Diabetes Metab Syndr Obes 2016 Nov;9:381-390 [FREE Full text] [doi: 10.2147/DMSO.S95247] [Medline: 27853384]

Abbreviations

GDM: gestational diabetes mellitus
mHealth: mobile health

Edited by A Kushniruk; submitted 08.02.22; peer-reviewed by S Schembre, H Ranjani, E Stringer, H Mehdizadeh; comments to author 28.06.22; revised version received 09.09.22; accepted 10.09.22; published 12.10.22.

Please cite as:
Kytö M, Koivusalo S, Ruonala A, Strömberg L, Tuomonen H, Heinonen S, Jacucci G
Behavior Change App for Self-management of Gestational Diabetes: Design and Evaluation of Desirable Features
JMIR Hum Factors 2022;9(4):e36987
URL: https://humanfactors.jmir.org/2022/4/e36987
doi:10.2196/36987
PMID:36222806

Original Paper

The Virtual Inclusive Digital Health Intervention Design to Promote Health Equity (iDesign) Framework for Atrial Fibrillation: Co-design and Development Study

Nino Isakadze[1,2], MD, MHS; Nancy Molello[3], MSB; Zane MacFarlane[1], BA; Yumin Gao[1], MsC; Erin M Spaulding[4,5], RN, PhD; Yvonne Commodore Mensah[3,4,5,6], RN, MHS, PhD; Francoise A Marvel[1,2], MD; Shireen Khoury[1,7], MD, MPH; Joseph E Marine[1], MD; Erin D Michos[1], MD, MHS; David Spragg[1], MD; Ronald D Berger[1], MD, PhD; Hugh Calkins[1], MD; Lisa A Cooper[3,4,6,8], MD, MPH; Seth S Martin[1,2], MD, MHS

[1]Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[3]Center for Health Equity, Johns Hopkins University, Baltimore, MD, United States

[4]School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]School of Nursing, Johns Hopkins University, Baltimore, MD, United States

[6]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD, United States

[7]Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, OH, United States

[8]School of Medicine, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Nino Isakadze, MD, MHS
Ciccarone Center for the Prevention of Cardiovascular Disease
Division of Cardiology, Department of Medicine
Johns Hopkins University School of Medicine
600 N Wolfe St/Halsted 500
Baltimore, MD, 21287
United States
Phone: 1 410 955 5999
Email: nisakad1@jh.edu

Abstract

Background: Smartphone ownership and mobile app use are steadily increasing in individuals of diverse racial and ethnic backgrounds living in the United States. Growing adoption of technology creates a perfect opportunity for digital health interventions to increase access to health care. To successfully implement digital health interventions and engage users, intervention development should be guided by user input, which is best achieved by the process of co-design. Digital health interventions co-designed with the active engagement of users have the potential to increase the uptake of guideline recommendations, which can reduce morbidity and mortality and advance health equity.

Objective: We aimed to co-design a digital health intervention for patients with atrial fibrillation, the most common cardiac arrhythmia, with patient, caregiver, and clinician feedback and to describe our approach to human-centered design for building digital health interventions.

Methods: We conducted virtual meetings with patients with atrial fibrillation (n=8), their caregivers, and clinicians (n=8). We used the following 7 steps in our co-design process: step 1, a virtual meeting focused on defining challenges and empathizing with problems that are faced in daily life by individuals with atrial fibrillation and clinicians; step 2, a virtual meeting focused on ideation and brainstorming the top challenges identified during the first meeting; step 3, individualized onboarding of patients with an existing minimally viable version of the atrial fibrillation app; step 4, virtual prototyping of the top 3 ideas generated during ideation; step 5, further ranking by the study investigators and engineers of the ideas that were generated during ideation but were not chosen as top-3 solutions to be prototyped in step 4; step 6, ongoing engineering work to incorporate top-priority features in the app; and step 7, obtaining further feedback from patients and testing the atrial fibrillation digital health intervention in a pilot clinical study.

Results: The top challenges identified by patients and caregivers included addressing risk factor modification, medication adherence, and guidance during atrial fibrillation episodes. Challenges identified by clinicians were complementary and included patient education, addressing modifiable atrial fibrillation risk factors, and remote atrial fibrillation episode management. Patients brainstormed more than 30 ideas to address the top challenges, and the clinicians generated more than 20 ideas. Ranking of the ideas informed several novel or modified features aligned with the Theory of Health Behavior Change, features that were geared toward risk factor modification; patient education; rhythm, symptom, and trigger correlation for remote atrial fibrillation management; and social support.

Conclusions: We co-designed an atrial fibrillation digital health intervention in partnership with patients, caregivers, and clinicians by virtually engaging in collaborative creation through the design process. We summarize our experience and describe a flexible approach to human-centered design for digital health intervention development that can guide innovative clinical investigators.

(JMIR Hum Factors 2022;9(4):e38048) doi:10.2196/38048

KEYWORDS

atrial fibrillation; digital health intervention; human-centered design; health equity; smartphone; mobile application; cardiac; cardiology; virtual meeting; virtual health; medication adherence; health equity

Introduction

Background

Digital health interventions (DHIs) are rapidly reshaping health care, including the delivery of cardiovascular care. Health and fitness smartphone app utilization is increasing, with 84 million users overall in the United States in 2021 [1]. Racial and ethnic minorities have higher rates of smartphone use for health-related purposes than White individuals in the United States [2,3]. This represents a unique opportunity to reach diverse groups of patients outside hospital walls and empower them with disease management knowledge and tools. By engaging patients in self-management, DHIs have the potential to improve access, management, and ultimately health outcomes of patients with cardiovascular disease. There are more than 50,000 health-management apps available in the Apple and Android Google Play app stores [4,5], but the majority do not adequately engage patients [6]. This problem is likely multifactorial, with causes that include a lack of evaluation of patient needs and development of products without grounding within established behavior change techniques. To design successful DHIs, innovators would benefit from identifying challenges related to management of the disease from patient, caregiver, and clinician perspectives in the early phases of app development. Furthermore, to ensure that the end products are useful to patients across different age, race, ethnicity, sex, and socioeconomic groups, engaging diverse individuals in the design process is crucial. This approach has been supported by a randomized study in which app selection based on target users' needs led to better usability and user satisfaction [7].

Human-centered design (HCD) is a core creative design methodology developed in the 1970s to facilitate innovative solution development processes. HCD methodology has been widely adopted in engineering and social sciences. More recently, HCD has been adopted for health care intervention development (eg, web-based health app co-design for refugee and migrant women, diabetes management app development, heart failure digital tool development, and personalized integrated care platform development for people with neurodegenerative disorders) [8-12]. To date, more than 100 design methodologies have been described that use different approaches to HCD, including participatory design and affinity diagramming, among other methodologies [9]. While there are differences in how various design methodologies are structured, many share the integral components of understanding end user needs through user engagement in the design process. DHIs developed with an HCD methodology have the potential to promote person-centered care, as they are designed in a way that respects patient needs, values, and preferences; allows care continuity outside the walls of the hospital; empowers patients with education; and provides emotional support. Overall, while general field guides for HCD are widely available, structured frameworks to guide innovators and investigators in designing disease management DHIs by applying HCD principles, including relevant stakeholders, and addressing the challenges in diverse patient engagement are limited. This provides an opportunity to describe our approach to applying HCD principles in the health care setting and contribute to the growing literature [13-15].

Case Study Significance

Atrial fibrillation (AF) is the most common cardiac arrhythmia. The lifetime risk of AF is 1 in 5 among Black individuals and 1 in 3 among White individuals [16,17]. AF is associated with poor quality of life [18], poor subjective health [18-20], and increased morbidity and mortality [21,22]. Scientific breakthroughs in therapy (including anticoagulation and rhythm control) and interventions (including advanced ablation techniques and left atrial appendage closure) have not been uniformly adopted in practice, especially among individuals from disenfranchised racial and ethnic groups [23-28]. In order for scientific advances to reach diverse groups of patients, there is a need to identify patient-, caregiver-, and clinician-level challenges and to engage all stakeholders in the process of creating innovative solutions. We previously developed a minimally viable AF DHI by tailoring the Corrie Health digital platform. This platform consists of the patient-facing Corrie app, connected devices, and a clinician dashboard, where actionable insights are stored to inform clinical decision-making. The Corrie Health DHI was shown to successfully improve outcomes among patients recovering from myocardial infarction

[29-31]. Several features of the original Corrie app were based on behavioral change theories, such as social cognitive theory and the health belief model, which includes domains of perceived susceptibility, perceived benefits, barriers, cues to action, self-efficacy, enactive attainment, knowledge, and outcome expectations [32]. Behavior change techniques are an important addition to the patient feedback processes described above, as they promote patient engagement and uptake of DHIs.

We aimed to further develop the AF DHI on the Corrie Health platform to create a mature product that reflects the needs of diverse patients, caregivers, and clinicians. Furthermore, we aimed to summarize our findings from the process and to propose an approach to HCD tailored for DHI development in an inclusive manner: the Virtual Inclusive Digital Health Intervention Design to Promote Health Equity (iDesign) framework. This framework can be applied to the development and optimization of other DHIs with relevant stakeholder input. Our goal is to help guide clinical investigators and health app developers to create engaging, equitable, and scalable digital health solutions.

Methods

Study Setting

To conduct this qualitative research study, we recruited an expert designer from the Johns Hopkins Center for Health Equity (author NM) who had completed in-person training in HCD methodology. The expert designer led the design process along with a clinician (author NI) who had experience in caring for patients with AF and had also completed online training in HCD methodology. We also recruited 4 individuals to help capture the meeting discussions in real time, including 2 clinicians, 1 undergraduate student, and 1 postgraduate student.

Ethics Approval

The study was approved by the Johns Hopkins Institutional Review Board (IRB00230621).

Patient Recruitment

We conducted purposeful sampling with the goal of having representation from multiple ethnic and racial backgrounds. While we were able to enroll White, Black, and Asian individuals, we were not successful in enrolling Hispanic or Native American individuals due to the very small number of eligible individuals in the source population. We recruited 8 patients with a diagnosis of AF and their caregivers or health partners (if the patient chose to include them) who were scheduled to see a cardiologist or electrophysiologist within 6 months of enrollment. All participants were recruited virtually via phone call, and consent was obtained by email. Design meetings were conducted on the Zoom platform. Virtual study procedures were chosen to allow for a safe study process in the setting of the COVID-19 pandemic. This also allowed us to increase access and engage eligible patients with a limited ability to attend in-person visits due to barriers such as work, household circumstances, and transportation [33].

Inclusion criteria were a diagnosis of AF, ownership of a smartphone (iPhone 5 or newer), residency in the United States, and proficiency in English.

Clinician Recruitment

Clinicians were recruited using flyers and emails with information regarding the study. Eligible clinicians included cardiologists, electrophysiologists, pharmacists, and advanced care providers. All clinicians had more than 3 years of experience caring for patients with AF. All participants provided informed consent virtually via email, as discussed above. Methods for each step of the iDesign process are outlined in the following sections and are also summarized in Figure 1.

Figure 1. Steps of the Virtual Inclusive Digital Health Intervention Design to Promote Health Equity (iDesign) framework.

We strived to be as inclusive as possible throughout the process, within the limitations of the COVID-19 pandemic, by (1) including a health equity expert (who was also an expert in HCD) to co-lead the HCD sessions, (2) inviting patients, caregivers, and clinicians with different backgrounds and expertise to drive the development of the components of the AF DHI, (3) offering virtual study visits that allowed for the inclusion of individuals who would not be able to take time off from work or travel to in-person visits, (4) offering one-to-one technology onboarding to allow the participation of individuals with different levels of technology literacy, and (5) ensuring that everyone contributed throughout the meetings by providing enough time for participants to speak and to reflect at the end of the session on whether their ideas were correctly captured. While the study team carefully considered providing a loaner phone with a data plan to patients who did not own that technology or have internet access, we decided at the time that device recirculation may have posed an undue risk of COVID-19 transmission.

Step 1: Defining Challenges and Empathizing

Patients

We conducted a virtual 1.5-hour meeting using audio-video conferencing on the Zoom platform with patients and caregivers or health partners. Discussions during the meeting were guided by the prespecified agenda. After introductions and an ice-breaking activity, the study leads explained to the participants that the goal of the session was for the study investigators to understand challenges patients experience while living with AF and to understand their journey after the diagnosis. Patients were also educated about the HCD approach and were reminded of the study goal to develop a DHI that would address the challenges of diverse patients with AF. Patients were divided into 2 groups to allow them to express their experiences with minimal time constraints. Participants were encouraged to think about either their personal experience as a patient or as a caregiver supporting a patient. We explained to participants that we would be using a journey map to capture their experiences following their AF diagnosis and their life with AF. We used the software Lucid (Lucid Software, Inc) to capture these experiences, which were projected on the shared Zoom screen. When the participants shared their experiences, we asked them to mark their timeline with notable milestones, which could be appointments, months, special holidays, or periods that marked how they were feeling. The study team used the following prompt during the breakout sessions to develop the journey map: “Let’s take a moment to think back to your first interaction with the healthcare team you had when diagnosed with AF. We want to hear your story as well as your partner’s story. Walk us through your journey after the diagnosis of AF.” We encouraged participants to focus on the following questions while answering the prompt: “What were your thoughts?” “How did you feel?” “What went well?” and “What were your challenges?”

We then shared the journey maps and the overall broad themes that emerged from the discussion with the whole group. The study participants were asked if their thoughts were adequately captured. Afterward, we thanked participants for their time and reminded them about the next steps, which included identifying top challenges that we would address together, followed by ideation, prototyping, and testing solutions.

Although we had identified broad themes during the sessions, we also conducted a qualitative analysis of a transcript of the audio recording of the session using the software ATLAS.ti (ATLAS.ti GmbH). First, each sentence was inductively coded via descriptive codes that were then clustered into broader codes sharing similar themes. These were further clustered into broad, emerging themes and formed into insight statements (these are discussed in the results section). Coding was conducted by the clinician-researcher (NI) and reviewed by an expert designer (NM) to ensure the dependability of the process.

Clinicians

We condensed step 1 (defining challenges) and step 2 (ideation) for the clinician meeting in order to accommodate the study participants’ schedules. We conducted a 1.5-hour meeting virtually via Zoom and recorded the session. Like the patient and caregiver meeting, the clinician meeting was guided by the prespecified agenda. After introductions and an ice-breaking activity, we asked the clinicians to share their responses to the following open-ended questions: (1) “What challenges do you face while caring for patients with AF?” and (2) “What are some patient strategies that increase their success in AF management?”

We summarized clinician responses using virtual whiteboarding and affinity diagraming. Affinity diagraming includes grouping together conceptually similar notes captured during discussion and identifying broad, emerging themes. Broad themes that were identified from the clinician sessions were then transformed into insight statements.

Step 2: Ideating Solutions (Patients Only)

Following step 1, a qualitative analysis of the meeting script was used to identify broad themes, which were converted into insight statements, as outlined above. We then constructed “how might we” questions from the insight statements. “How might we” questions are core components of the HCD approach and are meant to promote creativity [34]. We conducted a 1.5-hour meeting via Zoom with the goal of brainstorming and ideating solutions for the top challenges patients and caregivers expressed during the first meeting. We asked participants if the list of “how might we” questions generated from the transformation of emerging themes reflected what they shared during the first meeting. Participants were given the freedom to add or modify any questions. After confirming the accuracy of the “how might we” question list, we asked everyone to vote on the top-4 priority questions they wanted to brainstorm about.

Following the guidance of our expert human-centered designer, we decided to deploy the existing technology after the ideation phase. This was done to avoid bias in the thought process and facilitate generation of creative ideas without exposing participants to the existing technology. Additionally, we felt it was necessary for participants to have an understanding of existing capabilities and overall app design prior to engaging in prototyping.

Step 3: Individualized Virtual Onboarding (Patients Only)

We set up a Zoom meeting at a convenient time for each study participant to conduct individualized virtual onboarding to the minimally viable version of the Corrie app, which we named Corrie Afib. We followed a systematic approach for onboarding participants, which is described in Multimedia Appendix 1. The time required for onboarding each participant varied considerably, ranging from 15 minutes to 2 hours, based on the participant's digital health literacy. Patients attended this session without caregivers.

Step 4: Virtual Prototyping (Patients Only)

During the third 1.5-hour Zoom meeting, we first asked the study participants to identify the top solutions they wanted to design from the list of solutions that were generated during the second meeting (discussed in the results section). Study participants then voted using the Zoom chat box on the top solutions they wanted to prototype. The participants were split into 3 groups, and each group was given the task of prototyping the intervention. Virtual prototyping was conducted using Google Jamboard, which lets study designers capture patients' design ideas in real time and allows them to confirm their accuracy by continuously asking questions.

Step 5: Further Ranking

Additional ranking with study investigators and engineers was conducted, because the patients expressed their desire for other solutions to be incorporated in the DHI, although due to the time constraints of the virtual meetings, prototyping of only 3 solutions was feasible. Three study coinvestigator-clinicians and engineers met for discussion and ranked more than 30 patient ideas and more than 20 clinician ideas (Multimedia Appendices 2 and 3) that were not prototyped by patients; the ranking was based on feasibility, viability, and desirability and used a 3-point Likert scale [35]. Top-scoring solutions were prioritized for incorporation into the app, in addition to the ideas chosen by patients for prototyping. Our approach, like the HCD process, was flexible. When feasible, the ranking process can be combined with the ideation session. We then leveraged the Theory of Health Behavior Change as a conceptual model to guide further incorporation of the features in the design process [36,37]. The Theory of Health Behavior Change considers 3 major components driving individual behavior change: knowledge and beliefs, self-regulation skill and ability, and social facilitation (Multimedia Appendix 4) [36,37].

Step 6: Engineering Work (Ongoing)

Once specific functionalities were prioritized, the engineering team began to integrate the selected features into the app.

Step 7: Testing (Future)

The app functionalities informed by our iDesign process will be modified iteratively until no further major feedback is obtained from pilot clinical testing in a larger group of patients (N=100), who will be either Android or iOS users.

Results

Step 1: Defining Challenges and Empathizing

Patients

Patients (N=8) and their caregivers or health partners (N=3) joined the first group discussion virtually over the Zoom platform. Subsequent sessions were attended only by patients, without caregivers or health partners. The median age of the patients was 69 (IQR 58-78) years. There were 2 Black participants, 1 Asian participant, and 4 women; 4 patients had commercial insurance and 4 patients had Medicare insurance.

After conducting a qualitative analysis, we identified the following insight statements: (1) addressing risk factor modification and medication adherence is important; (2) guidance during AF episodes is important, because the episodes cause anxiety; (3) patients seek more information and education regarding AF, particularly regarding life after AF diagnosis; (4) patients seek education on medications and need help with medication adherence; (5) patients need to understand therapy options; and (6) patients need peer support. We then converted these insight statements into the following "how might we" questions (Textbox 1).

Textbox 1. List of "how might we" questions.

"How might we" questions (patients)
1. "HMW [how might we] use the Afib [atrial fibrillation] app to guide me in making lifestyle changes?"
2. "HMW use the Afib app to better guide me when I have symptoms of Afib?"
3. "HMW use the Afib app for education after diagnosis of Afib?"
4. "HMW use the Afib app to manage my medications?"
5. "HMW use the Afib app to help me better understand different options of care when living with Afib?"
6. "HMW use the Afib app for community support?"
"How might we" questions (clinicians)
7. "HMW better educate patients using the Afib app?"
8. "HMW better communicate with patients using the Afib app?"

Clinicians

We recruited 8 clinicians (including 2 women), although 2 of the recruited clinicians were not available for the meeting due to a schedule conflict. As described above, we combined defining the problems (ie, empathizing) and ideating solutions into 1 meeting for the clinicians. We identified the following insight statements through affinity diagraming during the session: (1) patient education is inadequate; (2) addressing AF modifiable risk factors is a priority; and (3) remote AF management and communication are needed.

We then constructed "how might we" questions (Textbox 1) based on these insight statements. Notably, insight statements from patient and caregiver input were similar to the ones generated by clinicians.

Step 2: Ideating Solutions

Patients

Patients selected the top 3 questions from Textbox 1 that they felt were most desirable to address. The top questions selected were questions 1, 2, and 5. The broad range of creative ideas brainstormed by the patients, numbering more than 30, are outlined in Multimedia Appendix 2.

Clinicians

The broad range of creative ideas brainstormed by clinicians, numbering more than 20, are outlined in Multimedia Appendix 3.

Step 3. Individualized Virtual Onboarding (Patients Only)

After one-on-one onboarding sessions with patients, we identified several learning points to further streamline the process for future studies. First, prior to connecting with the patient via Zoom, the study team member should send an introductory email explaining to the patient what they need prior to onboarding, including (1) their Apple ID and password or Google Play ID and password, (2) a stable Wi-Fi connection, and (3) a smartphone or, ideally, another device (such as a laptop, iPad, or computer) where they can watch an explanation of the onboarding process by the study team member while downloading the app to their phones and completing onboarding. Second, there is a need for 2 versions of onboarding videos, one for individuals who are more comfortable using technology and the other for those who need more guidance, to streamline the process while still addressing the needs of diverse individuals. The video directed toward individuals who are comfortable with technology should be brief and should only show each feature and give one example of using that feature, while the version directed toward individuals with low digital literacy should review the details of how to open an email from the study team and download the app, retrieve the passwords needed for app installation, set up an automated login (if desired), and give more than one example of using each feature of the app. Third, the study team should be available for questions and technical support.

Step 4: Virtual Prototyping (Patients Only)

At the end of the ideation meeting, we asked patients to vote on the top 3 ideas that they felt were most desirable to prototype. Table 1 shows the voting results.

Participants were then split into 3 groups, and the groups were given the task of prototyping each feature. Models of working prototypes were created with patient input for (1) providing information about AF triggers (Multimedia Appendix 5), (2) symptom and rhythm correlation for alerting the clinician team and family while experiencing AF (Multimedia Appendix 6), and (3) improved medication tracking (Multimedia Appendix 7).

Table 1. Patient choices on what to prototype and forming "what to prototype" statements.

What to prototype?	Patient choices from ideation session
AF[a] triggers	• Include information about AF triggers.
How to know you are having AF when you are or are not having symptoms; how the app can guide you during an episode	• Have a way to know if you are having AF when you do not recognize symptoms, such as by providing heart rate and blood pressure. Have a feature to learn if a fast heart rate is from AF or something else. • Provide the ability to connect with clinicians, friends, or family via the app when you have symptoms, to see if they can help with the decision on what to do next. • Quicker connection to people, especially during night hours, to help make decisions about next steps.
Improved medication tracking	• Include information on side effects of medications and what different options are available. • Include information on the effects of medications on lifestyle. • Indicate what medication and dose to take, and provide a reminder to take the medication at the correct time. • Include pictures of the different medications to help recognize them.

[a]AF: atrial fibrilation.

Step 5: Further Ranking

A second round of ranking identified patient education, peer support, and clinician engagement in remote AF management as viable, feasible, and desirable functionalities to incorporate within the app. New or modified functionalities chosen to be incorporated were aligned with the components of the Theory of Health Behavior Change (Multimedia Appendix 4).

Steps 6 and 7: Engineering Work and Testing

Engineering and development are in progress to expand the feature scope of the Corrie Afib app and the clinician dashboard based on HCD. Features include an expanded library for tailored AF education, risk factor tracking, rhythm monitoring for potential triggers, symptom and rhythm correlation for better therapeutic guidance, and community support (Multimedia Appendix 4).

Timeline for each step of the design process is included in Multimedia Appendix 8. Even after completion, this iterative process will continue with feedback obtained from patients enrolled in a pilot study of 100 participants.

Discussion

Principal Findings and Literature Review

In this exploratory co-design project, patients, their caregivers, and clinicians identified salient challenges in AF care, including inadequate education and guidance on AF management topics, such as lifestyle changes and guideline-directed medical therapy, inadequate resources to guide decision-making during an AF episode, and poorly structured patient-clinician communication during AF episodes. Similar qualitative and co-design experiences have not yet been described for AF management, but the challenges identified by our study participants overlap with known gaps in AF care outlined in the literature. In line with our participants' identification of inadequate education and guidance on AF management, there is evidence that a significant proportion of individuals with AF have inadequate disease-specific health literacy [26]. Furthermore, anticoagulation therapy, a core component of AF management to prevent stroke, is only implemented in less than half of individuals who are at high risk for stroke [25]. Significant disparities in stroke prevention therapy also exist, as women and Black individuals receive anticoagulation therapy less often than men and White individuals [24,27]. In addition, Black and low-income individuals are less likely to receive rhythm control therapy, including catheter ablation [23]. Educating and empowering patients regarding these core principles of AF management, as suggested by the patients and clinicians, has the potential to improve implementation of guideline-directed therapies. We co-designed and coprototyped solutions together with patients, caregivers, and clinicians and identified top features that will inform further development of AF DHIs to address the identified challenges. Further feasibility of our AF DHI will be assessed in a pilot study and, subsequently, in a definitive clinical study to assess its efficacy in improving quality of life and cardiovascular outcomes.

To date, HCD concepts have been applied in the design of DHIs for management of diseases including diabetes, heart failure, and metabolic syndrome [10,38,39]. A company specializing in HCD was asked to design an app for Ascensia Diabetes Care Holdings AG with the goal of educating and monitoring patients with diabetes; however, a detailed description of the methods is not available [34]. Joshi et al [38] engaged 10 dietetic interns to define problems, ideate, and prototype a mobile app for patients with metabolic syndrome. While clinicians were engaged in the design process, patient-level input, which is critical to meeting the needs of end users, was missing. Zachari et al [39] designed a diabetes self-management app using participatory design, in which understanding patient challenges through empathy was at the center of the design process; however, clinician involvement was missing, as was the case for the design of a heart failure digital tool [10]. Ahmed et al [11] described an investigation targeting the need for development of an integrated care platform for individuals with neurodegenerative disease, incorporating feedback from patients, clinicians, and other stakeholders with a mixed methods approach; however, the ideation component of HCD was not reported in that study. Given the limited literature on using comprehensive methodologies to apply HCD principles specifically to DHI design, we used this paper to summarize our experience in developing an AF DHI and to outline our design framework, iDesign, as an approach to HCD (Figure 1). The iDesign framework is based on core HCD principles and can be used as a template for investigators planning to develop or optimize DHIs. Throughout the co-design process, we have made a number of findings that will help promote an inclusive design process and streamline the iDesign framework for future innovative projects (Multimedia Appendix 9). Future study teams may build on or modify individual components of the iDesign process as deemed appropriate for particular scenarios. The iDesign framework may serve as an overall guide for the process of DHI development for disease management.

Limitations

First, while our study engaged a diverse group of patients, caregivers, and clinicians in the co-design process, our sample size was modest, and we may not have captured insights from participants who do not speak English or have uncommon backgrounds and experiences. However, our study included participants from various backgrounds, and both the patients and clinicians shared similar challenges and experiences, suggesting that we attained inductive thematic saturation. Furthermore, clinicians with years of experience caring for patients with a rich array of backgrounds were engaged as stakeholders, which may have counterbalanced this limitation. In a subsequent stage of testing the AF DHI, we will recruit a larger sample size with broader representation to solicit further feedback. Second, while the virtual co-design approach facilitated inclusiveness by overcoming barriers such as transportation or work schedules, a virtual environment may not be ideal for participants who have limited technology skills or do not have access to technology or the internet. By engaging patients with various levels of comfort with technology in our design process, we learned the significance of the onboarding process, in which skills can be taught to help overcome the technology-literacy challenge. To overcome a lack of access to technology and barriers to internet access, we plan to work with

local communities, as well as policy makers, insurance agencies, and other stakeholders to facilitate both design and further implementation of the DHI.

Conclusions

We identified key recommendations from patients, caregivers, and clinicians with diverse backgrounds that informed the development of a DHI for AF management on the Corrie Health digital platform. We have extended our design process experience and formulated a structured inclusive framework, iDesign, which stems from components of HCD, for informing the development and optimization of DHIs. This framework adds to the body of knowledge on approaches to HCD, is flexible, and is transferable to the development and adaptation of DHIs.

Acknowledgments

This study was supported by grants from the American Heart Association, including Health Technology and Innovation Strategically Focused Research Networks (SFRN) grants (20SFRN35380046, 20SFRN35490003) and a Community Health in Action grant to Johns Hopkins University. In addition, we are grateful for funding from Sandra and Larry Small and the Pollin Digital Innovation Fund. We thank Ariel Hicks, our designer, who created the figure for this paper.

Conflicts of Interest

Under a license agreement between Corrie Health and Johns Hopkins University, the university owns equity in Corrie Health. The university, FAM, and SSM are entitled to royalty distributions related to technology described in this study. Additionally, FAM and SSM are founders of and hold equity in Corrie Health. This arrangement has been reviewed and approved by Johns Hopkins University in accordance with its conflict of interest policies. FAM and SSM have also received research and material support from Apple and iHealth. EMS serves as a consultant to Corrie Health. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Steps of the onboarding process to the Corrie Afib application.
[DOCX File , 14 KB - humanfactors_v9i4e38048_app1.docx]

Multimedia Appendix 2
Responses to “How Might We” questions during the patient ideation session.
[DOCX File , 17 KB - humanfactors_v9i4e38048_app2.docx]

Multimedia Appendix 3
Responses to “How Might We” questions during the clinician ideation session.
[DOCX File , 15 KB - humanfactors_v9i4e38048_app3.docx]

Multimedia Appendix 4
Theoretical basis of behavior change strategies in Corrie Afib.
[DOCX File , 18 KB - humanfactors_v9i4e38048_app4.docx]

Multimedia Appendix 5
"What triggers Atrial fibrillation" feature prototype design by patient input.
[DOCX File , 144 KB - humanfactors_v9i4e38048_app5.docx]

Multimedia Appendix 6
“Way to know you are having Atrial fibrillation when you are or aren’t having symptoms and how the application can guide you during an episode” feature design prototype.
[DOCX File , 110 KB - humanfactors_v9i4e38048_app6.docx]

Multimedia Appendix 7
“Improve medication adherence” feature design prototype.
[DOCX File , 302 KB - humanfactors_v9i4e38048_app7.docx]

Multimedia Appendix 8
Time Required for each step of the design process.
[DOCX File , 13 KB - humanfactors_v9i4e38048_app8.docx]

Multimedia Appendix 9
Guidance on using Virtual (i)nclusive Digital Health Intervention Design to Promote Health Equity (iDesign) Methodology.
[DOCX File , 24 KB - humanfactors_v9i4e38048_app9.docx]

References

1. Ceci L. Revenue of health and fitness apps in the United States from 1st quarter 2017 to 4th quarter 2020, by app store. Statista. URL: https://www.statista.com/statistics/1226387/heath-fitness-app-revenue-us-quarter/ [accessed 2022-10-04]
2. Cooper LA. Why are health disparities everyone's problem?. Baltimore, MD: Johns Hopkins University Press; 2021.
3. Anderson M. Racial and ethnic differences in how people use mobile technology. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/ [accessed 2022-10-04]
4. Ceci L. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 2nd quarter 2022. Statista. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2022-10-04]
5. Ceci L. Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 2nd quarter 2022. Statista. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2022-10-04]
6. Singh K. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications. Issue Brief (Commonw Fund). URL: https://www.issuelab.org/resources/25066/25066.pdf [accessed 2022-09-23]
7. Lee J, Kim J. Can menstrual health apps selected based on users' needs change health-related factors? A double-blind randomized controlled trial. J Am Med Inform Assoc 2019 Jul 01;26(7):655-666 [FREE Full text] [doi: 10.1093/jamia/ocz019] [Medline: 30946478]
8. Catalani C, Green E, Owiti P, Keny A, Diero L, Yeung A, et al. A clinical decision support system for integrating tuberculosis and HIV care in Kenya: a human-centered design approach. PLoS One 2014;9(8):e103205 [FREE Full text] [doi: 10.1371/journal.pone.0103205] [Medline: 25170939]
9. Martin B, Hanington BM. Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Beverly, MA: Rockport Publishers; 2014:222-224.
10. Johnston W, Keogh A, Dickson J, Leslie SJ, Megyesi P, Connolly R, et al. Human-centered design of a digital health tool to promote effective self-care in patients with heart failure: mixed methods study. JMIR Form Res 2022 May 10;6(5):e34257 [FREE Full text] [doi: 10.2196/34257] [Medline: 35536632]
11. Ahmed M, Marín M, Bouça-Machado R, How D, Judica E, Tropea P, et al. Investigating users' and other stakeholders' needs in the development of a personalized integrated care platform (procare4life) for older people with dementia or parkinson disease: protocol for a mixed methods study. JMIR Res Protoc 2021 Jan 12;10(1):e22463 [FREE Full text] [doi: 10.2196/22463] [Medline: 33433394]
12. Human-Centred Approach for Health. UNICEF. URL: https://www.hcd4health.org/ [accessed 2022-09-23]
13. Bartlett R, Boyle JA, Simons Smith J, Khan N, Robinson T, Ramaswamy R. Evaluating human-centred design for public health: a case study on developing a healthcare app with refugee communities. Res Involv Engagem 2021 May 30;7(1):32 [FREE Full text] [doi: 10.1186/s40900-021-00273-2] [Medline: 34053451]
14. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: A scoping review of applications and contexts. PLoS One 2017;12(11):e0186744 [FREE Full text] [doi: 10.1371/journal.pone.0186744] [Medline: 29091935]
15. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
16. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines and the Heart Rhythm Society in collaboration with the Society of Thoracic Surgeons. Circulation 2019 Jul 09;140(2):e125-e151 [FREE Full text] [doi: 10.1161/CIR.0000000000000665] [Medline: 30686041]
17. Mou L, Norby FL, Chen LY, O'Neal WT, Lewis TT, Loehr LR, et al. Lifetime risk of atrial fibrillation by race and socioeconomic status: ARIC Study (atherosclerosis risk in communities). Circ Arrhythm Electrophysiol 2018 Jul;11(7):e006350 [FREE Full text] [doi: 10.1161/CIRCEP.118.006350] [Medline: 30002066]
18. Zhang L, Gallagher R, Neubeck L. Health-related quality of life in atrial fibrillation patients over 65 years: A review. Eur J Prev Cardiol 2015 Aug;22(8):987-1002. [doi: 10.1177/2047487314538855] [Medline: 24924742]
19. Rienstra M, Lubitz SA, Mahida S, Magnani JW, Fontes JD, Sinner MF, et al. Symptoms and functional status of patients with atrial fibrillation: state of the art and future research opportunities. Circulation 2012 Jun 12;125(23):2933-2943 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.069450] [Medline: 22689930]
20. Rienstra M, Lyass A, Murabito JM, Magnani JW, Lubitz SA, Massaro JM, et al. Reciprocal relations between physical disability, subjective health, and atrial fibrillation: the Framingham Heart Study. Am Heart J 2013 Jul;166(1):171-178 [FREE Full text] [doi: 10.1016/j.ahj.2013.02.025] [Medline: 23816037]

21. Ganesan AN, Chew DP, Hartshorne T, Selvanayagam JB, Aylward PE, Sanders P, et al. The impact of atrial fibrillation type on the risk of thromboembolism, mortality, and bleeding: a systematic review and meta-analysis. Eur Heart J 2016 May 21;37(20):1591-1602. [doi: 10.1093/eurheartj/ehw007] [Medline: 26888184]
22. Kim MH, Johnston SS, Chu B, Dalal MR, Schulman KL. Estimation of total incremental health care costs in patients with atrial fibrillation in the United States. Circ Cardiovasc Qual Outcomes 2011 May;4(3):313-320. [doi: 10.1161/CIRCOUTCOMES.110.958165] [Medline: 21540439]
23. Eberly LA, Garg L, Yang L, Markman TM, Nathan AS, Eneanya ND, et al. Racial/ethnic and socioeconomic disparities in management of incident paroxysmal atrial fibrillation. JAMA Netw Open 2021 Feb 01;4(2):e210247 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0247] [Medline: 33635328]
24. Essien UR, Holmes DN, Jackson LR, Fonarow GC, Mahaffey KW, Reiffel JA, et al. Association of race/ethnicity with oral anticoagulant use in patients with atrial fibrillation: findings from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation II. JAMA Cardiol 2018 Dec 01;3(12):1174-1182 [FREE Full text] [doi: 10.1001/jamacardio.2018.3945] [Medline: 30484833]
25. Hsu JC, Maddox TM, Kennedy KF, Katz DF, Marzec LN, Lubitz SA, et al. Oral anticoagulant therapy prescription in patients with atrial fibrillation across the spectrum of stroke risk: insights from the NCDR PINNACLE Registry. JAMA Cardiol 2016 Apr 01;1(1):55-62. [doi: 10.1001/jamacardio.2015.0374] [Medline: 27437655]
26. Reading SR, Go AS, Fang MC, Singer DE, Liu IA, Black MH, Anticoagulation and Risk Factors in Atrial Fibrillation–Cardiovascular Research Network (ATRIA‐CVRN) Investigators. Health literacy and awareness of atrial fibrillation. J Am Heart Assoc 2017 Apr 11;6(4):e005128 [FREE Full text] [doi: 10.1161/JAHA.116.005128] [Medline: 28400367]
27. Thompson LE, Maddox TM, Lei L, Grunwald GK, Bradley SM, Peterson PN, et al. Sex differences in the use of oral anticoagulants for atrial fibrillation: a report from the National Cardiovascular Data Registry (NCDR) PINNACLE Registry. J Am Heart Assoc 2017 Jul 19;6(7):e005801 [FREE Full text] [doi: 10.1161/JAHA.117.005801] [Medline: 28724655]
28. Tsao C, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
29. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
30. Marvel FA, Spaulding EM, Lee MA, Yang WE, Demo R, Ding J, et al. Digital health intervention in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2021 Jul;14(7):e007741 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.121.007741] [Medline: 34261332]
31. Spaulding EM, Marvel FA, Lee MA, Yang WE, Demo R, Wang J, et al. Corrie Health digital platform for self-management in secondary prevention after acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2019 May;12(5):e005509 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.119.005509] [Medline: 31043065]
32. 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] [Medline: 23512568]
33. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Randomized trial of a virtual cardiac rehabilitation program delivered at a distance via the Internet. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):952-959. [doi: 10.1161/CIRCOUTCOMES.114.001230] [Medline: 25271050]
34. The Field Guide to Human-Centered Design. IDEO.ORG. URL: https://www.designkit.org/resources/1 [accessed 2022-10-04]
35. Aifah A, Okeke NL, Rentrope CR, Schexnayder J, Bloomfield GS, Bosworth H, et al. Use of a human-centered design approach to adapt a nurse-led cardiovascular disease prevention intervention in HIV clinics. Prog Cardiovasc Dis 2020;63(2):92-100 [FREE Full text] [doi: 10.1016/j.pcad.2020.02.013] [Medline: 32092444]
36. Ryan P. Integrated theory of health behavior change: background and intervention development. Clin Nurse Spec 2009;23(3):161-70; quiz 171 [FREE Full text] [doi: 10.1097/NUR.0b013e3181a42373] [Medline: 19395894]
37. Laranjo L. Social media and health behavior change. In: Inyed-Abdul S, Gabarron E, Lau A, editors. Participatory Health Through Social Media. Cambridge, UK: Academic Press; 2016.
38. Joshi A, Amadi C, Schumer H, Galitzdorfer L, Gaba A. A human centered approach to design a diet app for patients with metabolic syndrome. Mhealth 2019;5:43 [FREE Full text] [doi: 10.21037/mhealth.2019.08.13] [Medline: 31620470]
39. Zachary WW, Michlig G, Kaplan A, Nguyen N, Quinn CC, Surkan PJ. Participatory design of a social networking app to support type ii diabetes self-management in low-income minority communities. Proc Int Symp Hum Factors Ergon Healthc 2017 Jun;6(1):37-43 [FREE Full text] [doi: 10.1177/2327857917061010] [Medline: 31157286]

Abbreviations

AF: atrial fibrillation
DHI: digital health intervention

HCD: human-centered design
iDesign: Virtual Inclusive Digital Health Intervention Design to Promote Health Equity

Edited by A Kushniruk; submitted 16.03.22; peer-reviewed by I Iyamu, J Lee; comments to author 06.04.22; revised version received 28.06.22; accepted 02.08.22; published 31.10.22.

Please cite as:
Isakadze N, Molello N, MacFarlane Z, Gao Y, Spaulding EM, Commodore Mensah Y, Marvel FA, Khoury S, Marine JE, Michos ED, Spragg D, Berger RD, Calkins H, Cooper LA, Martin SS
The Virtual Inclusive Digital Health Intervention Design to Promote Health Equity (iDesign) Framework for Atrial Fibrillation: Co-design and Development Study
JMIR Hum Factors 2022;9(4):e38048
URL: https://humanfactors.jmir.org/2022/4/e38048
doi:10.2196/38048
PMID:36315217

Original Paper

Adapting a Patient-Reported Outcome Measure to Digital Outpatient Specialist Health Care Services for Type 1 Diabetes: User Involvement Study

Heidi Holmen[1,2], PhD; Tone Singstad[3], MSc; Lis Ribu[4], PhD; Annesofie Lunde Jensen[5,6,7], PhD; Nina Mickelson Weldingh[8], MSc; Astrid Torbjørnsen[1], PhD

[1]Department of Nursing and Health Promotion, Oslo Metropolitan University, Oslo, Norway
[2]The Intervention Centre, Oslo University Hospital, Oslo, Norway
[3]Endocrinology Outpatient Service, Akershus University Hospital, Lørenskog, Norway
[4]Centre for Senior Citizen Staff, Oslo Metropolitan University, Oslo, Norway
[5]Department of Clinical Medicine, Faculty of Health, Aarhus University, Aarhus, Denmark
[6]Steno Diabetes Centre Aarhus, Aarhus University Hospital, Aarhus, Denmark
[7]ResCenPI - Research Centre for Patient Involvement, Aarhus University & the Central Denmark Region, Aarhus, Denmark
[8]Department of Research Support Service, Akershus University Hospital, Lørenskog, Norway

Corresponding Author:
Heidi Holmen, PhD
Department of Nursing and Health Promotion
Oslo Metropolitan University
Pilestredet 32
Oslo, 0130
Norway
Phone: 47 67235000
Email: heidi.holmen@oslomet.no

Abstract

Background: Diabetes self-management is crucial for patients with type 1 diabetes, and digital services can support their self-management and facilitate flexible follow-up. The potential of using digital patient-reported outcome (PRO) measures in routine outpatient care is not fully used owing to a lack of adapted PRO measures.

Objective: This study presents the process of identifying and adapting a digital PRO measure for use in clinical diabetes practice and describes the preferred item topics of the adapted PRO measure, as reported by patients and diabetes specialist nurses.

Methods: With the involvement of patients, diabetes specialist nurses, management, and researchers, we hosted a series of workshops and 2 dialogue conferences. Scoping searches to identify relevant PRO measures formed the foundation for the process. An in-person dialogue conference was conducted with diabetes specialist nurses as participants, and a digital dialogue conference was conducted with patients with type 1 diabetes as participants. A diabetes-specific PRO measure was translated and adapted to our digital platform. Notes and summaries from the dialogue conferences were imported into NVivo (QSR International) and thematically analyzed as a single combined data set.

Results: The thematic analysis of the 2 dialogue conferences aimed to explore the views of patients with type 1 diabetes and diabetes specialist nurses on the outcomes necessary to measure. An overarching theme, *Ensuring that the PRO measure captures the patients' needs precisely and accurately, in a way that facilitates care and communication with health care personnel*, was identified and supported with data from both the patients and diabetes specialist nurses. This theme contained four categories: *The need for explanatory text after questions to ensure understanding and accurate response*, *Capturing individual needs in standardized questions, getting to the heart of the patient's problem*, and *The questions increase patient reflection*.

Conclusions: We successfully conducted an iterative process that identified a PRO measure aligned with the topics raised by the diabetes specialist nurses. Similarly, the patients found the PRO measure to be relevant and one that was addressing their needs. Only minor adjustments were necessary when programming the PRO measure in the digital platform. Our management, patients, and diabetes specialist nurses had a valuable impact on the results. User involvement facilitated a specific focus on the

clinical requests to be met by PRO measures and how they must be adapted to local and digital platforms. Overall, this has facilitated the current implementation of the adapted digital PRO measure.

(JMIR Hum Factors 2022;9(4):e38678) doi:10.2196/38678

KEYWORDS

patient-reported outcome measures; user involvement; type 1 diabetes; digital interventions

Introduction

Background

Approximately 10% of the 463 million people with diabetes have type 1 [1], placing an increasing strain on limited resources in the health care service. Diabetes self-management is crucial to live a healthy life, attending to the following treatment cornerstones: what one eats, how one exercises, and medication and blood glucose to prevent complications [2]. In addition to microvascular and macrovascular complications, diabetes can be associated with psychological problems such as diabetes distress [3], anxiety [4], and depression [5]. Although self-management is a continuous affair [6], the knowledge needed to self-manage can increase stress. For example, as the patients become more knowledgeable, their fear of late complications and long-term consequences may compromise quality of life [7]. Altogether, the complex demands of diabetes are a burden on the patient and health care services. Patients with type 1 diabetes are usually offered diabetes follow-up organized through diabetes teams in hospital outpatient services [8,9]. Attitudes toward diabetes care have evolved from compliance thinking toward the support of patients active in their diabetes self-management [2]. In addition, the advancement of medical equipment for insulin delivery and glucose management has evolved with insulin pumps and continuous glucose monitoring systems, which forms new demands to the delivery of health services. However, the organization of health services, with fixed consultations at suboptimal or inconvenient times for patients, often hinders this evolution in routine care [2]. Failure to attend scheduled outpatient appointments is an increasing problem among patients with type 1 diabetes and is associated with both suboptimal diabetes care and nonsustainable economic loss [10]. Combining the focus on objective clinical parameters such as hemoglobin A1c (HbA_{1c}) level, time in range, and similar parameters with subjective experiences of the symptom and treatment burden of patients could reduce the psychosocial strain of type 1 diabetes [2-5,7].

A subjective or self-reported measure is often labeled as *patient-reported outcome* (PRO), defined as "any report of the patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [11]. PRO measures are measures or methods used to gather these patient reports. This is complementary to objective measures such as blood pressure, height, weight, or various blood measures. The use of PRO measures is suggested to support communication between patients and health care personnel (HCP) and to individualize the provided health care, for instance, in an outpatient consultation [12-14].

Despite the recent clinical uptake of PRO measures, there is a paucity of systematic reviews of PRO measures in diabetes care, and the evidence base largely comprises primary studies. Studies investigating the use of PRO measures to guide outpatient consultations for patients with type 1 diabetes indicates both its usefulness for guiding consultations [15] and its perceived relevance [16]. In addition, from the perspective of HCP, organizational aspects such as time, training, and facilities of a preconsultation PRO measure must be accounted for in any clinical use of PRO measures [17]. Furthermore, a generic PRO system called AmbuFlex has been a pioneer in flexible follow-up and the use of PRO measures to assess the need for consultations [18,19]. Patients in AmbuFlex are sent a digital PRO measure 2 weeks before a scheduled consultation to assess their prioritized need for follow-up. The timing, duration, and content of consultations can be tailored to the individual needs and priorities of the patients. The goal is to get patients engaged in their own self-management, treatment, and need for health care as a support for chronic illness self-management [20]. HCP using AmbuFlex refer to both advantages and disadvantages of digital PRO measures [21]. AmbuFlex has recently been evaluated for patients with type 1 diabetes (DiabetesFlex version 1 containing 45 items), where it maintained safe diabetes management, improved well-being, and decreased the need for face-to-face visits [22,23]. Complementary to these findings, an analysis of preconsultation PRO measures shows that HCP, on one hand, need training to interpret patient responses [24]. In contrast, patients value the opportunity to add free-text details to their answers [24]. Furthermore, the subjective PRO measures must complement the objective clinical measures of diabetes self-management [25]. PRO measures in routine diabetes visits can be feasible and acceptable among patients and HCP [26].

Thus, the increased use of PRO measures in clinical practice can facilitate user involvement; however, research is scarce [12]. A recent review highlighted the importance of involving users and other stakeholders in the development of PRO measures and subsequent implementation and transformation of clinical services [14]. It adds to previous literature emphasizing how both patients' and HCP's preferences for relevant measures should be considered [27]. There is more to a life with diabetes than HbA_{1c}, and psychosocial outcomes are crucial as they are often associated with glycemic control [28]. Pragmatic studies investigating opportunities for the development and implementation of more user-oriented health services have real potential to improve services [14,23].

Implementation of PRO measures in routine care for type 1 diabetes complies well with the overarching responsibility for diabetes self-management within complex diabetes care. Health services need to move toward a service built upon the needs of and requests from patients and not those from the health service

to enhance the support of patient self-management. Implementing PRO measures can facilitate this change, and if they are developed and tailored together with users, we believe that actual change can be seen.

Aims and Objectives

We present the process of identifying and adapting a digital PRO measure for use in clinical diabetes practice and describe the preferred item topics of the adapted PRO measure, as reported by patients and diabetes specialist nurses.

Methods

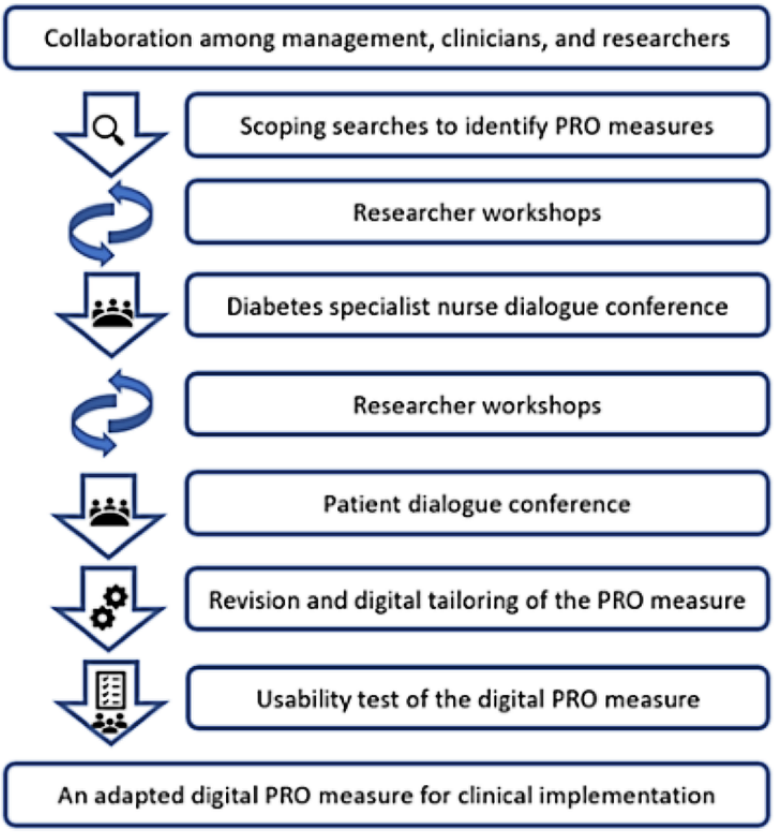
Dialogue Conference Methodology

Dialogue conferences are useful for sharing and gathering reflections that will have consequences for the invited participants. The suggested aim of dialogue conferences is to act as an arena for communication, where a broad range of participants can be mobilized [29]. In our study, the relevant participants were patients, diabetes specialist nurses, management, and researchers. Our intention was to conduct a series of dialogue conferences with all participants at the same time. Owing to the COVID-19 pandemic, we adapted our methodological process and conducted both in-person and digital conferences. Thus, the participants were invited on separate occasions; diabetes specialist nurses were invited to an in-person session, and patients were invited to a separate digital session. The digital sessions were hosted on Whereby (Whereby) [30], the platform used for treatment consultations and thus familiar to the patients. The research team and management of this study worked iteratively with the material throughout the study. The overall aim of the dialogue conferences was to enable all users to be active partners to facilitate the adaptation of a digital PRO measure in their health service.

Overall Design

This study was founded in a collaboration among management, clinicians, and researchers, with a common aim of implementing a digital PRO measure with clinical relevance for patients with type 1 diabetes at a specialized outpatient service. The stepwise methodological outline for this study is presented in Figure 1.

Figure 1. Process to identify, evaluate and adapt a set of patient-reported outcome measures for digital clinical implementation.



Participants

In total, four groups of users were included in the current process: management, researchers, diabetes specialist nurses, and patients. All invited participants accepted their invitation.

Management included 2 individuals, 1 (50%) trained as a diabetes specialist nurse holding a leadership position and 1 (50%) trained as a nurse holding a strategic position in organization development. Both had a Master's degree.

The researchers comprised a group affiliated with a university and a group affiliated with a university hospital in full-time or part-time clinical positions. All but one researcher were women (4/5, 80%). Of the 5 researchers, 2 (40%) were formerly trained as nurses, 1 (20%) was formerly trained as a diabetes specialist nurse, and 2 (40%) were physicians. All of them (5/5, 100%) had a clinical background in their field and a PhD.

At the outpatient clinic, 5 diabetes specialist nurses are employed. All (5/5, 100%) were invited to participate in the

dialogue conference, but only 80% (4/5) of them could attend. Thus, 4 female diabetes nurses with a range of 5 to 15 years of experience with patients with diabetes participated in the study.

Patient participants were recruited through convenience sampling and a personal invitation by a diabetes specialist nurse at the outpatient clinic. The patients comprised a group of 8 adults (n=2, 25% men and n=6, 75% women), with age ranging from early twenties to late seventies. Their ethnic backgrounds were from Europe and Asia. Their duration of diabetes ranged from 6 to 60 years, and they had a variety of experiences with treatment regimes, including multiple daily injections of insulin, daily injections of long-acting insulins, and insulin infusion pumps with a tube and those without a tube (such as the tubeless OmniPod). For self-monitoring of blood glucose, the participants shared their experiences with finger prick and various systems for continuous glucose monitoring and flash glucose monitoring.

Scoping Searches to Identify Relevant PRO Measures

To identify relevant PRO measures, we conducted systematic and exploratory searches to obtain an overview [31]. An example of the search is available in Multimedia Appendix 1. Overall, two researchers (HH and AT) independently assessed the citations and included those of relevance before they were discussed with the research team. Data extraction was conducted according to an a priori set of criteria including author and year of publication, name of the instrument, author and year of the development or validation of the instrument, aim of the instrument, number of items, availability of the PRO measure in Norwegian or another Scandinavian language, whether it had open access or was licensed, any experiences from use in clinical practice or research, and other comments regarding the instrument.

Analysis of the Dialogue Conferences

The dialogue conferences were not recorded, but notes were taken consecutively by 2 designated researchers at the dialogue conference with the diabetes specialist nurses, and 3 designated researchers took notes at the dialogue conference with the patients. These notes were imported into NVivo (QSR International) for storage and analysis. Thematic analysis, as described by Braun and Clarke [32,33], was applied as it is a widely adopted method for qualitative analysis and suggested to be an accessible approach for researchers. It consists of six phases: (1) familiarizing yourself with your data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report [32]. Notes from the abovementioned dialogue conferences provided the data for thematic analysis. The initial categories were discussed by the research group until consensus was reached, and the joint discussion led to the overall theme reflecting the content and meaning of the categories.

Usability Testing of the Digital PRO Measure and Digital Platform

The patient participants, management, and researchers were granted access to a test platform in which the tailored digital PRO measure was programed, with the opportunity to test the usability of the software system and items. They were asked to evaluate the PRO measure items, content, readability, interface, and usability and to provide feedback through the chat function in the software program. These findings are presented descriptively in the *Results* section.

Ethics Approval

The institutional local data protection officer approved the project (19/06920). All participants received written and oral information regarding the project and its aim, and informed consent was obtained from all participants.

Results

Results From the Scoping Searches

We identified 68 unique PRO measures aiming to address diabetes-specific outcomes. In a research workshop, the list of 68 PRO measures was reduced to an initial shortlist of 10 (15%) and further reduced to 5 (7%) PRO measures relevant for type 1 diabetes [22,34-37]. The main reason for exclusion was PRO measures addressing outcomes irrelevant for type 1 diabetes. The 7% (5/68) of the shortlisted PRO measures are listed in Table 1 along with their characteristics. These were presented at the diabetes specialist nurses dialogue conference.

Table 1. Shortlist of PRO[a] measures (n=5; alphabetically ordered).

Study (author, year)	Country	Name of the PRO measure	Aim of the PRO measure	Number of items	Available languages	Licensed	Used in clinical practice
Jensen et al [22], 2020	Denmark	DiabetesFlex version 1	To assess psychosocial and physical symptoms and prepare and focus the conversation between the patient and health care personnel	Maximum=45 (items are responsive, and questions are reduced or added based on the responses)	Danish	No	Yes
Brod et al [35], 2006	Denmark	Diabetes Medication Satisfaction (Diab-Med-Sat)	To measure diabetes treatment satisfaction, applicable to a wide range of diabetes therapies	23	English, Danish, Norwegian, and many more	Yes	Yes
Gold et al [37], 1994	United Kingdom	Gold	To assess hypoglycemia awareness	1	English, Norwegian, and many more	No	Yes
Carlton et al [36], 2017	United Kingdom	Health and Self-management in Diabetes–10 (HASMID)	To measure the impact of self-management in diabetes	8	English	No	No
Alvarado-Martel et al [34], 2017	Spain	Vida con Diabetes tipo 1 (ViDa1)	To measure health-related quality of life in patients with type 1 diabetes	34	Spanish and English	No	No

[a]PRO: patient-reported outcome.

Results From the Initial Researcher Workshops

In preparation for the dialogue conferences and summary work after the conferences, a series of workshops was conducted among the management and researchers. In preparation for the first series of workshops, the findings from the scoping searches were systematized by HH and AT in a Microsoft Excel file containing the name of the instrument, author and year of publication, where the instrument was identified, whether it was used in Norway, availability in Norwegian, whether it was used in research or clinically, aim of the instrument, and who made the original development. The extracted data were sent to the research team. In total, 2 videoconference workshops were conducted, aiming to reduce the number of relevant PRO measures and form a shortlist of instruments before the diabetes specialist nurse dialogue conference. Between these 2 workshops, the researchers and management reviewed a list of identified diabetes-specific PRO measures and noted their comments on the PRO measure and its characteristics as extracted. Each member commented according to their individual thoughts of relevance as a preparation for joint discussion. A new series of workshops among the researchers was conducted to summarize the data gathered in the first dialogue conference and to prepare PRO measures to discuss with patients during a new dialogue conference.

Summary of Diabetes Specialist Nurses Dialogue Conference to Explore and Evaluate PRO Measures

The dialogue conference with the diabetes specialist nurses opened with the opportunity to freely discuss their views on diabetes care today, whether and how digital service may be useful, opportunities with electronic health records, and how these may provide inspiration for future services.

Post-it notes were used to scribble ideas for specific items or questions that they regarded as crucial to gather information on. The diabetes specialist nurses were concerned that the information gathered using a PRO measure is designed for self-reporting by the patient. Where patients are responsible to self-report their current situation and needs, the nurses were afraid that they would miss out on an opportunity to detect something that the patient would have left out from their self-report—intentionally or unintentionally. A nurse simply said, “we need the obvious,” without adding details to what the obvious are. They initially discussed details regarding glycemic control more than other aspects of diabetes; however, after some time, the discussion shifted. After discussing the details regarding hypoglycemic events, stability or lack thereof in HbA_{1c} level, treatment regime, technical equipment, injection technique, time in range, and any fluctuations in blood glucose that patients cannot explain, there seemed to be a gradual shift toward the more psychosocial aspects of diabetes. The diabetes specialist nurses emphasized that there is more to life than diabetes, and, sometimes, the treatment regime should be adjusted because there is just too much going on. Obtaining this knowledge through items with fixed responses may be challenging, and free text was suggested as valuable.

When the initial open discussion ended, a short break was taken before the 5 PRO measures presented in Table 1 were discussed. Overall, the diabetes specialist nurses reported that all the

presented PRO measures had some relevant items. They were familiar with the 1-item PRO measure addressing hypoglycemia in plain language—"Do you know when your hypos are commencing?" answered on a 7-point Likert scale ranging from *1* (*always aware*) to *7* (*never aware*), developed by Gold et al [37], and found it useful. However, in addition to this item on hypoglycemia awareness, the overall impression was of the negative framing of the items, asking about how troublesome or problematic the aspects of diabetes treatment were. Having diabetes is obviously life-changing, but not something that can be removed. Asking about what is normal was also perceived as problematic by the diabetes specialist nurses because *normal* can mean many things. Furthermore, they questioned the relevance of items on mood and discussed the importance of asking patients about parts of their treatment that they could offer some guidance on. The majority of the diabetes specialist nurses were in favor of the diabetes-specific questions. Overall, items on hypoglycemia, support, control, and social activities were considered as important. The DiabetesFlex questionnaire version 1 [22] covered much of what had already been discussed as necessary and was perceived as relevant and more positively framed; however, some cultural adaptation from Danish to Norwegian would increase its relevance.

Results From the Researcher Workshops Between the Dialogue Conferences

After the nurses' dialogue conference, the researchers met to summarize and plan for the next step. Relevant topics, themes, and items were collected based on the input and notes from the diabetes specialist nurses, and 56 item topics were identified (Multimedia Appendix 1). These item topics were presented to the management and researchers in a workshop, and as a result we found that Diabetes Flex [22] covered most of these item topics in a favorable manner. The decision to go forth with the Diabetes Flex was presented and discussed in the clinic, and mutual agreement was reached among the researchers and clinical staff to concentrate on the items in the DiabetesFlex. Being a result of many years of collaboration among researchers, patients, and HCP [22,23], it was particularly regarded as clinically relevant. At this point, researcher ALJ was not participating in these discussions or the decision to move forward with a Norwegian adaptation of DiabetesFlex; rather, she was consulted afterward. As the research team behind DiabetesFlex had revised version 1 (45 items) and version 2 was available, we concentrated on the latter with 69 items. The overlap with our identified items and desired framing of questions served as a confirmation of relevance (Multimedia Appendix 1).

The authors, TS and NMW, drafted the first translation of the DiabetesFlex questionnaire version 2, and AT, LR, and HH reviewed and revised the translation. Items already translated into Norwegian, such as the World Health Organization–5 (WHO-5), were identified and used. The final translation of the remaining items that were not previously translated was discussed with researchers and clinicians before it was presented to ALJ, who participated in the Danish development and testing of the DiabetesFlex.

Summary of Patient Dialogue Conference to Explore and Evaluate PRO Measures

The Norwegian translation of the DiabetesFlex questionnaire was distributed to patients who had signed up for a digital dialogue conference. The patients received two files: one file with the mandatory annual consultation questionnaire containing 66 items and one with the optional consultation questionnaire containing 37 items. There is an overlap between the 2 questionnaires (Multimedia Appendix 1). The digital dialogue conference was hosted with 8 patients, 4 members of the research team, and 1 observer from the learning and mastery center at the hospital. In short, the patients felt that the items covered their needs and were adequately framed and easy to understand, but they were numerous and seemed somewhat too much to handle at first glance. The patients also added some reflections regarding life in general. A patient said the following:

> *If life is hard at the moment, it is not necessarily a bad idea to discuss this.*

However, another patient suggested that diabetes should be the only focus of the questions.

Results From Thematic Analysis of Dialogue Conferences

Overview

In the following section, the combined thematic analysis of the 2 dialogue conferences is presented, along with a descriptive summary of the usability test of the adapted digital PRO measure. The aim of the 2 dialogue conferences was to explore the outcomes necessary to measure in type 1 diabetes and to assess a set of PRO measures. Both the diabetes specialist nurses and patients referred to some tacit knowledge when they discussed the relevance of questions, what to add, and what to delete. The thematic analysis [32] of the notes taken during the 2 conferences resulted in 4 categories (Textbox 1). These reflect the shared opinions among the patients and nurses on what a clinical PRO measure for diabetes must address to be relevant for all stakeholders. Through discussions among the researchers, an overall theme emerged from the 4 categories: *Ensuring that the PRO measure captures the patients' needs precisely and accurately, in a way that facilitates care and communication with HCP*.

The theme, *Ensuring that the PRO measure captures the patients' needs precisely and accurately, in a way that facilitates care and communication with HCP*, has four categories: *Need for explanatory text after questions to ensure understanding and accurate response*, *Capturing individual needs in standardized questions, getting to the heart of the patient's problem*, and *The questions increase patient reflection*, shedding light on the content and purpose as experienced by diabetes specialist nurses and patients with diabetes (Textbox 1).

Textbox 1. Theme and categories obtained from the thematic analysis of the dialogue conferences.

Theme

- Ensuring that the patient-reported outcome measure captures the patients' needs precisely and accurately, in a way that facilitates care and communication with health care personnel

Categories

- Need for explanatory text after questions to ensure understanding and accurate response
- Capturing individual needs in standardized questions
- Getting to the heart of the patient's problem
- The questions increase patient reflection

Need for Explanatory Text After Questions to Ensure Understanding and Accurate Response

Both the diabetes specialist nurses and patients highlighted the need for questions that were immediately understood and the need to explain in detail about what was asked for. An example could be to explain where the patients can find their HbA_{1c} level or what is meant by *time in range*. Explanatory text could also appear after the patient had given their response, to provide some immediate guidance to self-manage or link to help pages or other relevant sources. For example, if they indicated severe symptoms, such as chest pain or trouble breathing, text with information regarding emergency health services should appear, emphasizing the need to contact these services.

Hypoglycemia was raised as a frequent and important topic to address. However, when a patient feels a hypoglycemic episode commencing and the blood glucose values differ from one person to another. To address the burden of hypoglycemia, the patient participants suggested to simply ask if hypoglycemia was experienced as a problem, and thus to replace phrases such as "How often do you experience hypoglycemia with a blood glucose reading at 3.9 mmol/mol or lower?"

Some questions should have had some context added to their explanation, as they could be challenging to interpret and answer. Some patients had comorbidities that affected their health more than diabetes, and some clarification was requested, such as for questions asking to assess pain that should specify whether a response should be related to their general health or their health related to type 1 diabetes.

Capturing Individual Needs in Standardized Questions

The diabetes specialist nurses highlighted their need to apply a holistic care perspective. A nurse said the following:

> *We cannot help if we cannot see the whole person.*

Individual patients have unique needs, preferences, and beliefs, and a nurse mentioned an example of patients being on each side of a scale, where one may be totally ignorant of their disease, whereas the other is extremely frightened of late complications and fear that they will lose their eyesight tomorrow. Patients suggested that the opportunity to add their own text after some standardized questions could allow for their individual needs to be explained:

> *If you really want to know how we are doing, you have to add a free-text field.*

Another patient also suggested the need for free text related to situations in which the given response interval would not fit their response, for example, where they wanted to answer between 2 boxes.

Similarly, the diabetes specialist nurses discussed letting patients add information in free-text fields, but they were concerned about whether patients were able to recognize their needs and put these needs into words. They were skeptical about self-reporting as they had experiences with faulty self-reports, both intentionally by providing blood glucose readings that seemed better than they actually were or unintentionally, such as recall bias, where patients suggested that they had nocturnal hypoglycemic episodes, whereas it had only been the previous night.

Getting to the Heart of the Patient's Problem

The diabetes specialist nurses were not concerned about the number of questions. Especially, if there were boxes to tick or intervals to indicate, for example, units of insulin taken, the questionnaire would be rather quick to complete. They worried more about whether patients were able to indicate their problems by using a free-text field. Again, patients suggested that free-text fields had to be an option if HCP were to get to the heart of their problems. The patients felt that the questionnaire they had been presented was slightly lengthy, with 69 items, but they agreed that the topics were relevant and necessary. A patient also asked whether some of the questions could be responsive; that is, if you indicate having symptoms, a list will appear, but if you indicate no symptoms, you are spared from having to look at the list of symptoms.

Patients also acknowledged the questions asking about their *worry* or how often they think about diabetes-related problems or challenges. These topics were appreciated by the patients more than those asking about their quality of life. A patient said the following:

> *Some of these questions...being worried is one of the things that makes us tired. Very relevant question, this must be included. Very important question.*

The Questions Increase Patient Reflection

Patients stated through the dialogue conference that their thoughts had begun to wander as they read the questions they

had been sent in preparation. A patient said that the questions made her reflect upon her daily life and the choices she made. Another patient added some arguments and said that it had increased the motivation to obtain more knowledge and to obtain an HbA_{1c} level measurement more often. Furthermore, the questions regarding symptoms and late complications also prompted reflections among the patients, and, although it was brutal, a patient said the following:

> *This became a bit serious question, very important for reflection, to think through. This is the problematic side of the disease. Can stand exactly as it does.*

Similarly, the diabetes specialist nurses suggested that some topics were not necessarily important for them to know about, such as whether patients needed their prescriptions filled or whether they had attended their eye examination, but they could act as a reminder for the patients.

Usability Testing of the Digital PRO Measure and Platform

Patient participants (8/8, 100%), HCP (5/5, 100%), management (2/2, 100%), and researchers (5/5, 100%) were granted access to a test platform for usability testing. Input was accounted for in the process of tailoring the DiabetesFlex questionnaire to fit in a Norwegian digital platform. Responses were provided by 75% (6/8) of the invited patients who had participated in the previous dialogue conference, 60% (3/5) of the researchers, 100% (2/2) of the management representatives, and 100% (5/5) of the HCP. Responses included comments in favor of the free-text fields, that the items covered all necessary aspects of a life with type 1 diabetes, and that the inclusion of items covering areas they normally would not address in a traditional consultation was valued. More technical responses targeted the platform and covered feedback on items that obviously should not be mandatory, such as problems with erections that had to be answered by female respondents because of a default function in the digital system making all questions mandatory. Furthermore, if one missed a question, they would not automatically be sent back to that specific question; rather, they had to skim through all the questions to find the blank one and give their response.

Revisions of the PRO Measure

The PRO measure was revised according to feedback from the patients before the usability testing and after the usability testing, before clinical implementation. Most revisions were the addition of explanatory text to the items. For instance, in the first version, the patients were asked to report their HbA_{1c} level, and in the revised version, text was added to explain which value they should report: "What is your last HbA_{1c} (Last HbA_{1c} measured at any doctor's office within the last 6 weeks)." Similarly, when asking for the mean blood glucose level for the past 2 weeks, explanatory text was added to guide the patients to find and report the right value. Finally, we added an item asking the patients to specify the date on which the HbA_{1c} level was determined.

Discussion

Principal Findings

In this study, we intended to obtain deep insight into patients' and diabetes specialist nurses' requirements regarding PRO measures in diabetes type 1. Moreover, we aimed to explore how these requirements would transfer to a digital platform. Specifically, we wanted to identify and adapt a PRO measure that would meet the requirements and be suitable for digital reporting and interpretation.

The results of our iterative process led to the identification of an overall theme describing the demands of patients and diabetes specialist nurses: *Ensuring that the PRO measure captures the patients' needs precisely and accurately, in a way that facilitates care and communication with HCP*. The patients and diabetes specialist nurses showed great consistency when they spoke about their needs and demands. Both groups acknowledged the importance of glycemic control and the need to report details of blood glucose, hypoglycemia, and insulin regime. Both groups talked about the need to report how things really are, beyond numbers. Similarly, both groups mentioned problematic aspects of standardized questions, particularly relevant because diabetes type 1 can be very different from person to person [2-7]. Consistent with previous studies [15,17,24,25], our participants highlighted the importance of asking understandable questions, explaining them in detail if necessary, and adding open-text fields for patients to elaborate and aid an increased understanding of the patient's situation. Neither group reflected on the lack of age-adjusted PRO measures. However, making the items responsive of preceding reports were mentioned as favorable. This allows for some adjustments in the items the patient is presented with, based on their needs or characteristics.

The diabetes specialist nurses were concerned about the patients' ability to put their needs into written words. In contrast, the patients were concerned about not having the opportunity to use their own words to describe a problem. This highlights the challenge of how needs can be communicated through a format that is different from conversation. However, if a PRO measure report is outside the anticipated thresholds, a dialogue with HCP will have to occur and the patient will have an opportunity to elaborate [23]. However, both groups regarded PRO measures as relevant in prompting reflection among patients. Standardized items can address topics that the patient would not think about as relevant for their diabetes [24-26]. In this way, patients can become more knowledgeable about their disease, its signs and symptoms, and steps necessary for self-management through their digital self-report [20]. Although digital platforms for diabetes self-management support have been used for some time [38,39], there are few studies investigating how digital PRO measures can be used to assess and prioritize the need for contact with diabetes outpatient services [40,41]. Furthermore, the challenges associated with digital PRO measures in clinical care are the same in digital mode as when using paper. Our diabetes specialist nurses had some concerns about the risks of using digital PRO measures to assess and prioritize patients' need for follow-up. Thus, the PRO measures have to ask for the most relevant aspects of the disease, treatment, and

self-management [24-26]. In addition, the risk of self-report bias is not to be underestimated, both for overreporting and underreporting of symptoms [42]. Similarly, the threat of recall bias must be accounted for in the decisions on how frequently a PRO measure should be distributed to the patient. These factors can greatly influence the decisions made by HCP [24,26]. However, besides the risk of not asking for essential information, neither diabetes specialist nurses nor patients reflected on the practical interpretation of the patient report and the consequences the interpretation has for the patient treatment. The standardization of items may facilitate easy interpretation by the diabetes specialist nurses and explain why free text is seldom used in clinical PRO measures [43]. In addition, standardized questions may seem more relevant, prompting relevant disease actions among both HCP and patients, thus facilitating self-management and further use of digital PRO measures. Future studies should investigate how the PRO measures can prompt self-management and their associations with interpretations and treatment by HCP.

Most PRO measures that we identified were developed to act as outcomes in intervention studies and not to tailor clinical patient care. This lack of clinically relevant PRO measures that are easy to administer has been discussed previously [14]. Thus, the diabetes specialist nurses did not immediately see the clinical value of all the PRO measures. For example, the nurses worried that the PRO measure would reveal or address problematic areas in which they could not offer any treatment or follow-up. Furthermore, some of the PRO measures that we identified were found by the diabetes specialist nurses to be negatively framed, long, asking irrelevant questions, and not considering any subjective input (eg, through open-text fields). Digital platforms can, to an extent, offer more dynamic PRO measures that are tailored to the patient's characteristics, for example, their age, treatment, or challenges. However, more studies have to investigate the reliability of such dynamic alternations. With the current pandemic and potential shortage of HCP, new digital and dynamic methods to assess patient needs are warranted, and only few PRO measures relevant for clinical application in diabetes exist.

In our study, we aimed to adapt a PRO measure for digital, clinical use. When developing such new processes of care for implementation, both patient and HCP perspectives must be addressed [15,17,21,24,26]. Often, lack of personal motivation or engagement and low quality of the digital platforms are hindering implementation [44]. Furthermore, ease of use has been described as an important facilitator for digital interventions, and lack of exposure to or knowledge of digital interventions is considered as the most important barrier [45]. Through our early involvement of stakeholders from all levels, we have tried to prevent these barriers. Although user involvement can be time consuming and resource demanding, it is valuable as it addresses all stakeholder perspectives in the early development [25,45]. Our 2 dialogue conferences and the inclusion of users in the final usability test have contributed to the likelihood of implementing a relevant and useful PRO measure in a digital platform that is useful and easy to engage with [45]. Interestingly, we found that patient engagement was easy to pursue using a digital platform for communication and interaction, and the number of patients willing to participate was higher than that for our in-person meetings. The extent of this positive impact on recruitment should be investigated in future studies to facilitate increased user involvement.

Limitations

Our study has some limitations, particularly related to our user participants. We had no family members and other HCP besides diabetes specialist nurses, and we did not have the opportunity to gather all users in 1 meeting at the same time, which may have reduced the potential for valuable input, as we did not manage to facilitate joint discussions among use groups. However, we had a heterogeneous sample of patient representatives, ensuring a width in our input. Furthermore, we included the patient users later in the process than preferred. We acknowledge that early involvement of patients could have affected our results. We could have pursued the inclusion of more patient groups, but we were able to accomplish tailoring and adapting a PRO measure to a digital platform through valuable feedback. Our patients can be biased and more engaged in digital platforms than the average patient, and thus more capable of understanding the questions. During the recruitment of patients, we strived to obtain a heterogeneous group in terms of age, gender, and ethnicity. However, selection bias is a well-known challenge, and we acknowledge that our users can be a more well-functioning group, consistent with previous studies [24,25]. Our study was conducted in a Scandinavian country, adapting a Scandinavian PRO measure, and the PRO measures' relevance for other areas has not been addressed. Future studies should pursue adaptation in other countries to assess its relevance. Regarding the inclusion of other HCP, we consider diabetes specialist nurses as the most important group, as they are the first responders and those who are closest to patients through the current organizational structure. Thus, we believe that their engagement and response are of utmost importance, with specific impacts on the adapted PRO measure.

The chosen PRO measure contains a combination of previously validated measures and items, in addition to some self-developed ones. These have not been validated in combination, and we did not pursue to validate the adapted PRO measure in this study. This would be important to address in future studies.

Conclusions

We have shown how a process aiming to identify and adapt a PRO measure into a digital platform for clinical use in a diabetes outpatient clinic has been conducted. The involvement of management, patients, and diabetes specialist nurses had a valuable impact on our results. This process has been crucial in facilitating the forthcoming implementation success, as our stakeholders have contributed to a digital PRO measure that addresses relevant topics and is used in a user-friendly digital platform. Allowing user involvement using digital platforms was also found to be a favorable method, as it increased attendance beyond expectations. We anticipate positive outcome from our digital PRO measure because we ensured user involvement of highly invested stakeholders, thus ensuring the relevance of our PRO measure.

Acknowledgments

The authors would like to show their utmost appreciation to all the patients and diabetes specialist nurses who participated in this study, for their time and valuable insight.

Data Availability

The Norwegian version of the DiabetesFlex questionnaire can be made available upon request to ALJ. The English version 1 is published elsewhere [23].

Authors' Contributions

HH participated in the entire process, collected data, conducted the thematic analysis, and drafted and revised the manuscript. TS facilitated and participated in the entire process, recruited all the participants, and revised the manuscript draft. LR and NMW participated in the process and revised the manuscript draft. ALJ participated in the process, contributed valuable insight from DiabetesFlex and expertise regarding the clinical use of patient-reported outcome measures, and revised the manuscript. AT facilitated and participated in the entire process, collected data, and revised the manuscript. All authors read and approved the final manuscript. TS presented a poster based on this study at a national diabetes seminar.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Identified items and overlaps with the original DiabetesFlexTM questionnaire.
[DOCX File , 21 KB - humanfactors_v9i4e38678_app1.docx]

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, 9 edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
2. Hermanns N, Ehrmann D, Finke-Groene K, Kulzer B. Trends in diabetes self-management education: where are we coming from and where are we going? A narrative review. Diabet Med 2020 Mar;37(3):436-447. [doi: 10.1111/dme.14256] [Medline: 32017188]
3. Skinner TC, Joensen L, Parkin T. Twenty-five years of diabetes distress research. Diabet Med 2020 Mar;37(3):393-400. [doi: 10.1111/dme.14157] [Medline: 31638279]
4. Smith KJ, Béland M, Clyde M, Gariépy G, Pagé V, Badawi G, et al. Association of diabetes with anxiety: a systematic review and meta-analysis. J Psychosom Res 2013 Feb;74(2):89-99. [doi: 10.1016/j.jpsychores.2012.11.013] [Medline: 23332522]
5. Chireh B, Li M, D'Arcy C. Diabetes increases the risk of depression: a systematic review, meta-analysis and estimates of population attributable fractions based on prospective studies. Prev Med Rep 2019 Jun;14:100822 [FREE Full text] [doi: 10.1016/j.pmedr.2019.100822] [Medline: 30815337]
6. Thorpe CT, Fahey LE, Johnson H, Deshpande M, Thorpe JM, Fisher EB. Facilitating healthy coping in patients with diabetes: a systematic review. Diabetes Educ 2013;39(1):33-52 [FREE Full text] [doi: 10.1177/0145721712464400] [Medline: 23073967]
7. Speight J, Holmes-Truscott E, Hendrieckx C, Skovlund S, Cooke D. Assessing the impact of diabetes on quality of life: what have the past 25 years taught us? Diabet Med 2020 Mar;37(3):483-492. [doi: 10.1111/dme.14196] [Medline: 31797443]
8. American Diabetes Association Professional Practice Committee, Draznin B, Aroda VR, Bakris G, Benson G, Brown FM, et al. 4. Comprehensive medical evaluation and assessment of comorbidities: standards of medical care in diabetes-2022. Diabetes Care 2022 Jan 01;45(Suppl 1):S46-S59. [doi: 10.2337/dc22-S004] [Medline: 34964869]
9. 1. Diagnosis of diabetes, risk assessment and follow-up of people at high risk of developing diabetes. The Norwegian Directorate of Health. 2017. URL: https://www.helsedirektoratet.no/retningslinjer/diabetes [accessed 2022-10-23]
10. Brewster S, Bartholomew J, Holt RI, Price H. Non-attendance at diabetes outpatient appointments: a systematic review. Diabet Med 2020 Sep;37(9):1427-1442. [doi: 10.1111/dme.14241] [Medline: 31968127]
11. Santana MJ, Haverman L, Absolom K, Takeuchi E, Feeny D, Grootenhuis M, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. Qual Life Res 2015 Jul;24(7):1707-1718. [doi: 10.1007/s11136-014-0903-5] [Medline: 25589231]

12. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2018 Dec;2:42 [FREE Full text] [doi: 10.1186/s41687-018-0061-6] [Medline: 30294712]
13. de Wit M, Versloot J, Zenlea I, Goethals ER. Using person-reported outcomes (PROs) to motivate young people with diabetes. Curr Diab Rep 2020 May 16;20(7):23 [FREE Full text] [doi: 10.1007/s11892-020-01305-z] [Medline: 32415346]
14. Skovlund SE, Lichtenberg TH, Hessler D, Ejskjaer N. Can the routine use of patient-reported outcome measures improve the delivery of person-centered diabetes care? A review of recent developments and a case study. Curr Diab Rep 2019 Aug 16;19(9):84. [doi: 10.1007/s11892-019-1190-x] [Medline: 31420754]
15. Hernar I, Graue M, Strandberg RB, Lie SS, Sigurdardottir AK, Richards DA, et al. Young adults with type 1 diabetes and their experiences with diabetes follow-up and participation in the DiaPROM pilot trial: a qualitative study. Diabet Med 2021 Jun;38(6):e14535. [doi: 10.1111/dme.14535] [Medline: 33547702]
16. Hernar I, Graue M, Richards DA, Strandberg RB, Nilsen RM, Rekdal M, et al. Use of patient-reported outcome measures (PROMs) in clinical diabetes consultations: the DiaPROM randomised controlled pilot trial. BMJ Open 2021 Apr 14;11(4):e042353 [FREE Full text] [doi: 10.1136/bmjopen-2020-042353] [Medline: 33853796]
17. Haugstvedt A, Hernar I, Graue M, Strandberg RB, Stangeland Lie S, Sigurdardottir AK, et al. Nurses' and physicians' experiences with diabetes consultations and the use of dialogue tools in the DiaPROM pilot trial: a qualitative study. Diabet Med 2021 Jun;38(6):e14419. [doi: 10.1111/dme.14419] [Medline: 33035378]
18. Schougaard LM, Larsen LP, Jessen A, Sidenius P, Dorflinger L, de Thurah A, et al. AmbuFlex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. Qual Life Res 2016 Mar;25(3):525-534 [FREE Full text] [doi: 10.1007/s11136-015-1207-0] [Medline: 26790427]
19. Hjollund NH. Fifteen years' use of patient-reported outcome measures at the group and patient levels: trend analysis. J Med Internet Res 2019 Sep 30;21(9):e15856 [FREE Full text] [doi: 10.2196/15856] [Medline: 31573943]
20. Mejdahl CT, Schougaard LM, Hjollund NH, Riiskjær E, Thorne S, Lomborg K. PRO-based follow-up as a means of self-management support - an interpretive description of the patient perspective. J Patient Rep Outcomes 2017;2:38 [FREE Full text] [doi: 10.1186/s41687-018-0067-0] [Medline: 30238083]
21. Mejdahl CT, Schougaard LM, Hjollund NH, Riiskjær E, Lomborg K. Exploring organisational mechanisms in PRO-based follow-up in routine outpatient care - an interpretive description of the clinician perspective. BMC Health Serv Res 2018 Jul 13;18(1):546 [FREE Full text] [doi: 10.1186/s12913-018-3352-y] [Medline: 30001729]
22. Jensen A, Lomborg K, Hjollund N, Schougaard L, Olesen L, Laurberg T, et al. DiabetesFlex™ – the effect of PRO-based telehealth and user involvement in care management of patients with type 1 diabetes: trial protocol for a non-inferiority randomised controlled study. Medical Res Archiv 2020;8(7) [FREE Full text] [doi: 10.18103/mra.v8i7.2145]
23. Laurberg T, Schougaard LM, Hjollund NH, Lomborg KE, Hansen TK, Jensen AL. Randomized controlled study to evaluate the impact of flexible patient-controlled visits in people with type 1 diabetes: the DiabetesFlex trial. Diabet Med 2022 May;39(5):e14791. [doi: 10.1111/dme.14791] [Medline: 35028992]
24. Eeg-Olofsson K, Johansson U, Linder E, Leksell J. Patients' and health care professionals' perceptions of the potential of using the digital diabetes questionnaire to prepare for diabetes care meetings: qualitative focus group interview study. J Med Internet Res 2020 Aug 19;22(8):e17504 [FREE Full text] [doi: 10.2196/17504] [Medline: 32812884]
25. Skovlund SE, Troelsen LH, Klim L, Jakobsen PE, Ejskjaer N. The participatory development of a national core set of person-centred diabetes outcome constructs for use in routine diabetes care across healthcare sectors. Res Involv Engagem 2021 Sep 10;7(1):62 [FREE Full text] [doi: 10.1186/s40900-021-00309-7] [Medline: 34507618]
26. Skovlund SE, Troelsen LH, Noergaard LM, Pietraszek A, Jakobsen PE, Ejskjaer N. Feasibility and acceptability of a digital patient-reported outcome tool in routine outpatient diabetes care: mixed methods formative pilot study. JMIR Form Res 2021 Nov 03;5(11):e28329 [FREE Full text] [doi: 10.2196/28329] [Medline: 34730545]
27. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012 Oct;21(8):1305-1314. [doi: 10.1007/s11136-011-0054-x] [Medline: 22048932]
28. Glasgow RE, Peeples M, Skovlund SE. Where is the patient in diabetes performance measures? The case for including patient-centered and self-management measures. Diabetes Care 2008 May;31(5):1046-1050 [FREE Full text] [doi: 10.2337/dc07-1845] [Medline: 18445728]
29. Philips M, Huzzard T. Developmental magic? Two takes on a dialogue conference. J Organizational Change Manage 2007;20(1):8-25. [doi: 10.1108/09534810710715252]
30. Whereby homepage. Whereby. URL: https://whereby.com/) [accessed 2022-10-23]
31. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: scoping reviews (2020 version). JBI Manual for Evidence Synthesis, JBI, 2020. 2020. URL: https://jbi-global-wiki.refined.site/space/MANUAL/4687342/Chapter+11:+Scoping+reviews [accessed 2022-10-18]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]

34. Alvarado-Martel D, Ruiz Fernández MA, Cuadrado Vigaray M, Carrillo A, Boronat M, Expósito Montesdeoca A, et al. ViDa1: the development and validation of a new questionnaire for measuring health-related quality of life in patients with type 1 diabetes. Front Psychol 2017;8:904 [FREE Full text] [doi: 10.3389/fpsyg.2017.00904] [Medline: 28620331]
35. Brod M, Skovlund SE, Wittrup-Jensen KU. Measuring the impact of diabetes through patient report of treatment satisfaction, productivity and symptom experience. Qual Life Res 2006 Apr;15(3):481-491. [doi: 10.1007/s11136-005-1624-6] [Medline: 16547787]
36. Carlton J, Elliott J, Rowen D, Stevens K, Basarir H, Meadows K, et al. Developing a questionnaire to determine the impact of self-management in diabetes: giving people with diabetes a voice. Health Qual Life Outcomes 2017 Jul 18;15(1):146 [FREE Full text] [doi: 10.1186/s12955-017-0719-4] [Medline: 28720133]
37. Gold AE, MacLeod KM, Frier BM. Frequency of severe hypoglycemia in patients with type I diabetes with impaired awareness of hypoglycemia. Diabetes Care 1994 Jul;17(7):697-703. [doi: 10.2337/diacare.17.7.697] [Medline: 7924780]
38. Holmen H, Wahl AK, Cvancarova Småstuen M, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. J Med Internet Res 2017 Jun 23;19(6):e227 [FREE Full text] [doi: 10.2196/jmir.7045] [Medline: 28645890]
39. Fatehi F, Menon A, Bird D. Diabetes care in the digital era: a synoptic overview. Curr Diab Rep 2018 May 10;18(7):38. [doi: 10.1007/s11892-018-1013-5] [Medline: 29748905]
40. Gandrup J, Ali SM, McBeth J, van der Veer SN, Dixon WG. Remote symptom monitoring integrated into electronic health records: a systematic review. J Am Med Inform Assoc 2020 Nov 01;27(11):1752-1763 [FREE Full text] [doi: 10.1093/jamia/ocaa177] [Medline: 32968785]
41. Whear R, Thompson-Coon J, Rogers M, Abbott RA, Anderson L, Ukoumunne O, et al. Patient-initiated appointment systems for adults with chronic conditions in secondary care. Cochrane Database Syst Rev 2020 Apr 09;4:CD010763 [FREE Full text] [doi: 10.1002/14651858.CD010763.pub2] [Medline: 32271946]
42. Marciano L, Camerini A, Schulz PJ. The role of health literacy in diabetes knowledge, self-care, and glycemic control: a meta-analysis. J Gen Intern Med 2019 Jun;34(6):1007-1017 [FREE Full text] [doi: 10.1007/s11606-019-04832-y] [Medline: 30877457]
43. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014 Jun;23(6):508-518. [doi: 10.1136/bmjqs-2013-002524] [Medline: 24505110]
44. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
45. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
HCP: health care personnel
PRO: patient-reported outcome
WHO-5: World Health Organization–5

Edited by A Kushniruk; submitted 12.04.22; peer-reviewed by E Daskalaki, B Schreiweis, A Hornsten, J Leksell; comments to author 18.07.22; revised version received 12.09.22; accepted 06.10.22; published 15.11.22.

Please cite as:
Holmen H, Singstad T, Ribu L, Jensen AL, Weldingh NM, Torbjørnsen A
Adapting a Patient-Reported Outcome Measure to Digital Outpatient Specialist Health Care Services for Type 1 Diabetes: User Involvement Study
JMIR Hum Factors 2022;9(4):e38678
URL: https://humanfactors.jmir.org/2022/4/e38678
doi:10.2196/38678
PMID:36378513

Review

The Development and Use of Chatbots in Public Health: Scoping Review

Lee Wilson[1], PhD; Mariana Marasoiu[2], MPhil

[1]Centre for Policy Futures, University of Queensland, St Lucia, Queensland, Australia

[2]Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Lee Wilson, PhD
Centre for Policy Futures
University of Queensland
Level 3, Michie Building (9)
St Lucia, Queensland, 4072
Australia
Phone: 61 0795318198
Email: l.wilson7@uq.edu.au

Abstract

Background: Chatbots are computer programs that present a conversation-like interface through which people can access information and services. The COVID-19 pandemic has driven a substantial increase in the use of chatbots to support and complement traditional health care systems. However, despite the uptake in their use, evidence to support the development and deployment of chatbots in public health remains limited. Recent reviews have focused on the use of chatbots during the COVID-19 pandemic and the use of conversational agents in health care more generally. This paper complements this research and addresses a gap in the literature by assessing the breadth and scope of research evidence for the use of chatbots across the domain of public health.

Objective: This scoping review had 3 main objectives: (1) to identify the application domains in public health in which there is the most evidence for the development and use of chatbots; (2) to identify the types of chatbots that are being deployed in these domains; and (3) to ascertain the methods and methodologies by which chatbots are being evaluated in public health applications. This paper explored the implications for future research on the development and deployment of chatbots in public health in light of the analysis of the evidence for their use.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines for scoping reviews, relevant studies were identified through searches conducted in the MEDLINE, PubMed, Scopus, Cochrane Central Register of Controlled Trials, IEEE Xplore, ACM Digital Library, and Open Grey databases from mid-June to August 2021. Studies were included if they used or evaluated chatbots for the purpose of prevention or intervention and for which the evidence showed a demonstrable health impact.

Results: Of the 1506 studies identified, 32 were included in the review. The results show a substantial increase in the interest of chatbots in the past few years, shortly before the pandemic. Half (16/32, 50%) of the research evaluated chatbots applied to mental health or COVID-19. The studies suggest promise in the application of chatbots, especially to easily automated and repetitive tasks, but overall, the evidence for the efficacy of chatbots for prevention and intervention across all domains is limited at present.

Conclusions: More research is needed to fully understand the effectiveness of using chatbots in public health. Concerns with the clinical, legal, and ethical aspects of the use of chatbots for health care are well founded given the speed with which they have been adopted in practice. Future research on their use should address these concerns through the development of expertise and best practices specific to public health, including a greater focus on user experience.

(JMIR Hum Factors 2022;9(4):e35882) doi:10.2196/35882

KEYWORDS

chatbots; conversational agents; public health; evidence; scoping review; health care system; chatbot development; digital health; mental health; health technology; COVID-19; pandemic; chatbot application

Introduction

Sundar Pichai, the chief executive officer of Google, expressed in a recent interview his view that artificial intelligence (AI) will have a more profound impact on humanity than the advent of fire, the internet, or electricity [1]. Although Pichai has a vested interest in propagating visions of AI-enhanced futures, there is no doubting the extent to which advances in computing technology are driving rapid transformations in the ways in which we interact with computing systems, organizations, one another, and the world. A salient feature of this rapidly changing technological landscape is the burgeoning development and use of conversational agents, or "chatbots."

Chatbots—software programs designed to interact in human-like conversation—are being applied increasingly to many aspects of our daily lives. Made to mimic natural language conversations to facilitate interaction between humans and computers, they are also referred to as "conversational agents," "dialog assistants," or "intelligent virtual assistants," and they can support speech and text conversation. Notable early chatbots include ELIZA (1966;"a mock Rogerian psychotherapist"), PARRY (1972; a chatbot simulating a person with paranoid schizophrenia, developed by a psychiatrist in response to ELIZA), and ALICE (1995; a general conversational chatbot, inspired by ELIZA) [2]. Recent advances in the development and application of chatbot technologies and the rapid uptake of messenger platforms have fueled the explosion in chatbot use and development that has taken place since 2016 [3]. Improvements to natural language processing (NLP; which includes speech recognition, text-to-speech, speech-to-text, natural language understanding, and natural language generation), as well as the emergence and publicity of commercial "virtual assistants" such as Siri, Google Now, Cortana, and Alexa [4] have brought AI into many aspects of our daily lives. Chatbots are now found to be in use in business and e-commerce, customer service and support, financial services, law, education, government, and entertainment and increasingly across many aspects of health service provision [5].

The ongoing COVID-19 pandemic has further driven the rapid uptake and deployment of chatbots [6], many making use of commercial chatbot development platforms such as IBM's Watson Assistant, Google Dialogflow, Yellow Messenger, and Turn.io to develop custom chatbots to help combat the disease. In the face of the burden placed upon health care systems by the pandemic, chatbots have enabled the automation of services toward addressing the need for physical distancing and helped disseminate information and relieve the pressure on medical services by public health systems around the globe [7,8].

The use of AI for symptom checking and triage at scale has now become the norm throughout much of the world, signaling a move away from human-centered health care [9] in a remarkably short period of time. Recognizing the need to provide guidance in the field, the World Health Organization (WHO) has recently issued a set of guidelines for the ethics and principles of the use of AI in health [10]. WHO has itself made use of chatbots to provide guidance and combat misinformation about COVID-19 through its Health Alert chatbot [11] that communicates in a number of different languages through WhatsApp, Viber, and Facebook messenger, which has reportedly reached over 12 million people [12].

In the light of the huge growth in the deployment of chatbots to support public health provision, there is pressing need for research to help guide their strategic development and application [13]. This paper aimed to help address this deficit. We examined the evidence for the development and use of chatbots in public health to assess the current state of the field, the application domains in which chatbot uptake is the most prolific, and the ways in which chatbots are being evaluated. Reviewing current evidence, we identified some of the gaps in current knowledge and possible next steps for the development and use of chatbots for public health provision. Our research questions are as follows.

1. What does the evidence tell us about the use of chatbots in public health?
2. In which fields of public health have chatbots been used the most frequently?
3. What are the types of chatbots that have been used in public health?
4. How have chatbots in public health been evaluated?
5. What are the potential lessons to be learned from the evidence for the use of chatbots in public health?

Methods

We carried out a scoping review of studies on the use of chatbots in public health. We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines and methodological framework by Arksey and O'Malley [14] for scoping studies and searched the titles and abstracts of studies on the MEDLINE, PubMed, Scopus, Cochrane Central Register of Controlled Trials, IEEE Xplore, ACM Digital Library, and Open Grey databases over a period of 2 weeks in June 2021. Our search terms included "chatbot," "conversational agent," and their synonyms and public health, global public health, and related terms (see Multimedia Appendix 1). We chose to broaden our search to include health care and health to gain a broader understanding of the application domains in which chatbots are being used for health-related purposes. The domain categorization was assigned in 3 ways: (1) self-identified by the authors, (2) categorized according to current definitions of public health sectors, and (3) assigned according to the design scope of the chatbot. With regard to mental health, we made a further distinction between chatbots that were specifically designed to provide social support in nondiagnosed patients, defining these as "counseling/support," and those that were designed to deal with clinical illnesses such as depression, defining these as "mental health."

The use of AI and digital technologies and the roles in which they are deployed in health tend to blur the boundaries between population and clinical health—that is, chatbots that are used to service individual health needs are often equally as relevant to population-level health in application. In this respect, the synthesis between population-based prevention and clinical care

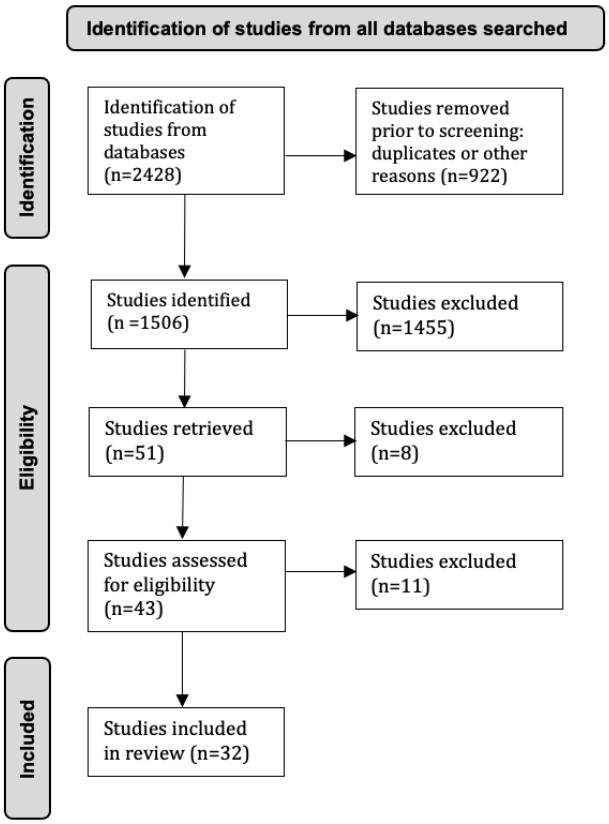
at an individual level [15] becomes particularly relevant. Implicit to digital technologies such as chatbots are the levels of efficiency and scale that open new possibilities for health care provision that can extend individual-level health care at a population level. We have therefore included studies of chatbots designed for the provision of health services to individuals where there is evidence of demonstrable health impacts and, importantly, where they have the potential for scalable efficiencies to support health outcomes at a population level.

Our selection methodology was as follows. One of the authors screened the titles and abstracts of the studies identified through the database search, selecting the studies deemed to match the eligibility criteria. The second author then screened 50% of the same set of identified studies at random to validate the first author's selection. The papers meeting the criteria for inclusion at the title- and abstract-screening stage were retrieved and reviewed independently by both authors, with subsequent discussion about discrepancies and resolution to end with an agreed upon list of included studies.

Our inclusion criteria were for the studies that used or evaluated chatbots for the purpose of prevention or intervention and for which the evidence showed a demonstrable health impact. We included experimental studies where chatbots were trialed and showed health impacts. We also included feasibility studies for agents that are being rolled out, randomized controlled trials (RCTs) informing the feasibility of conversational agents that have obvious applicability for scalability and potential for population-level interventions, and comparative analyses of in-service chatbots. We chose not to distinguish between embodied conversational agents and text-based agents, including both these modalities, as well as chatbots with cartoon-based interfaces.

We excluded thought experiments, design outlines and reflections on systems that have yet to be implemented, descriptions of proposed chatbots and conversational agents, prototypes of system architecture, surveys and predesign analyses, frameworks, commentaries, validation studies, technical papers that introduced agents explaining their architecture and design that have yet to be trialed, and papers exploring perceptions of digital agents or their acceptability or validity among users. We also excluded studies comparing the effect of differences in technical approaches (eg, messaging) and studies that used "Wizard of Oz" protocols—a protocol used to test users' reactions in which a human responds to users through an interface in which they think they are interacting with a computer. The review selection process is shown in Figure 1.

Figure 1. Review selection process.



Results

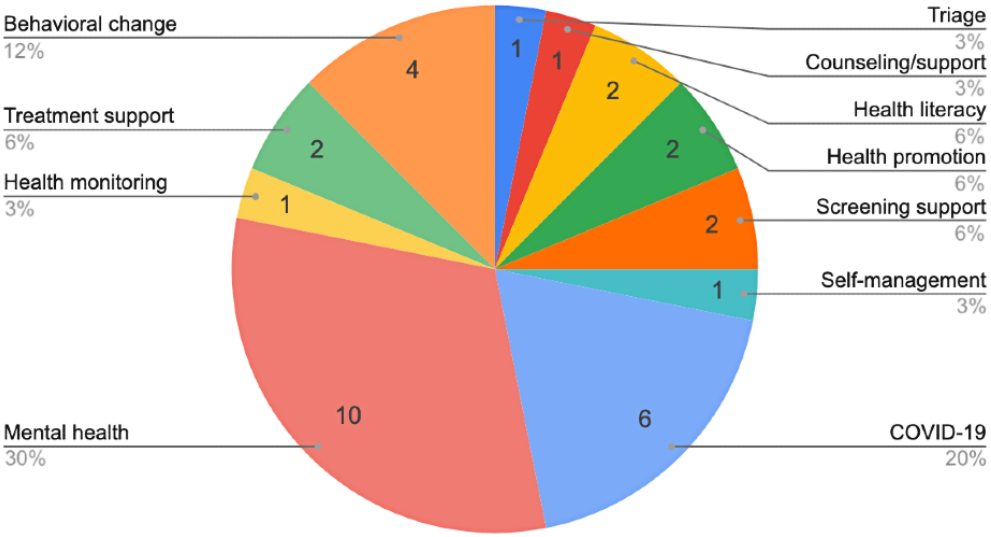
Included Studies

In total, 32 studies met the inclusion criteria. These studies included RCTs (n=12), user analytics (n=8), user experience studies (n=3), an experimental pilot (n=1), a descriptive study (n=1), comparative analyses (n=2), a case control study (n=1), design processes (n=2), and feasibility studies (n=2). These studies were distributed across 11 application domains.

Mental health and COVID-19 dominated the application domains. This result is possibly an artifact of the maturity of the research that has been conducted in mental health on the

use of chatbots and the massive surge in the use of chatbots to help combat COVID-19. The graph in Figure 2 thus reflects the maturity of research in the application domains and the presence of research in these domains rather than the quantity of studies that have been conducted.

Figure 2. Distribution of included publications across application domains. Mental health research and COVID-19 form the majority of the studies. Due to the small numbers of papers, percentages must be interpreted with caution and only indicate the presence of research in the area rather than an accurate distribution of research.

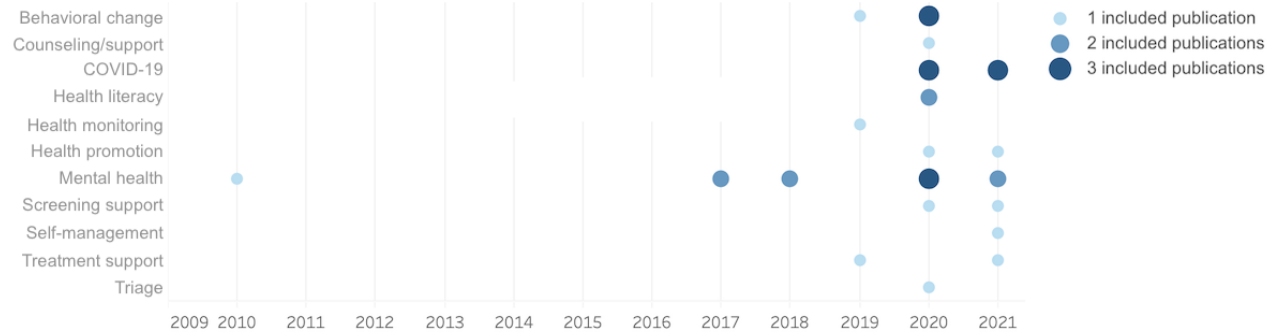


Maturity of Chatbot Research in Public Health Domains

The timeline for the studies, illustrated in Figure 3, is not surprising given the huge upsurge of interest in chatbots from 2016 onward. Although health services generally have lagged behind other sectors in the uptake and use of chatbots, there has been greater interest in application domains such as mental health since 2016. This finding may reflect both the degree to which conversational technologies lend themselves to the kinds of interactive methodologies used in mental health and the necessity for greater scrutiny of the methods that are used by health practitioners in field.

Similarly, one can see the rapid response to COVID-19 through the use of chatbots, reflecting both the practical requirements of using chatbots in triage and informational roles and the timeline of the pandemic.

Figure 3. Distribution of included publications across application domains and publication year. Mental health research has a continued interest over time, with COVID-19–related research showing strong recent interest as expected.



Chatbot Design

Studies that detailed any user-centered design methodology applied to the development of the chatbot were among the minority (3/32, 9%) [16-18]. Most (22/32, 69%) included papers only described broadly the messaging content available in the chatbot (eg, topics covered) and what functionality was available to the user (eg, daily reminders), but few had a description of the process by which those features and capabilities were decided upon.

One study that stands out is the work of Bonnevie and colleagues [16], who describe the development of *Layla*, a trusted source of information in contraception and sexual health among a population at higher risk of unintended pregnancy. *Layla* was designed and developed through community-based participatory research, where the community that would benefit from the chatbot also had a say in its design. *Layla* demonstrates the potential of AI to empower community-led health interventions. Such approaches also raise important questions about the production of knowledge, a concern that AI more broadly is undergoing a reckoning with [19].

Two-thirds (21/32, 66%) of the chatbots in the included studies were developed on custom-developed platforms on the web [6,16,20-26], for mobile devices [21,27-36], or personal

computers [37,38]. A smaller fraction (8/32, 25%) of chatbots were deployed on existing social media platforms such as Facebook Messenger, Telegram, or Slack [39-44]; using SMS text messaging [42,45]; or the Google Assistant platform [18] (see Figure 4).

All the included studies tested textual input chatbots, where the user is asked to type to send a message (free-text input) or select a short phrase from a list (single-choice selection input). Only 4 studies included chatbots that responded in speech [24,25,37,38]; all the other studies contained chatbots that responded in text.

The majority (28/32, 88%) of the studies contained very little description of the technical implementation of the chatbot, which made it difficult to classify the chatbots from this perspective. Most (19/32, 59%) of the included papers included screenshots of the user interface. However, some only provided sketches of the interface, and often, the text detailing chatbot capabilities was not congruent with the picture accompanying the text (eg, the chatbot was described as free entry but the screenshot showed a single-choice selection). In such cases, we marked the chatbot as using a combination of input methods (see Figure 5).

Surprisingly, there is no obvious correlation between application domains, chatbot purpose, and mode of communication (see Multimedia Appendix 2 [6,8,9,16-18,20-45]). Some studies did indicate that the use of natural language was not a necessity for a positive conversational user experience, especially for symptom-checking agents that are deployed to automate form filling [8,46]. In another study, however, not being able to converse naturally was seen as a negative aspect of interacting with a chatbot [20].

The presentation of the chatbot persona (see Figure 6) was usually presented as a static avatar (n=17). Of these chatbots, 8 were given an anthropomorphic avatar, whether as a photo or drawing (eg, clipart), whereas the rest adopted either a robot, animal, cartoon, or another abstract avatar. Embodied conversational agents (n=5) were only presented as “female” human-like avatars. Of those with no avatars (n=6), this absence was usually due to the platform restriction (eg, WhatsApp or some forms of embedded web chat). The influence of avatar presence and the anthropomorphic appearance of chatbots are still an underresearched area, but we expect it will be of particular importance for future chatbot design in health care.

Figure 4. Distribution of chatbot platforms in the included studies. PC: personal computer.

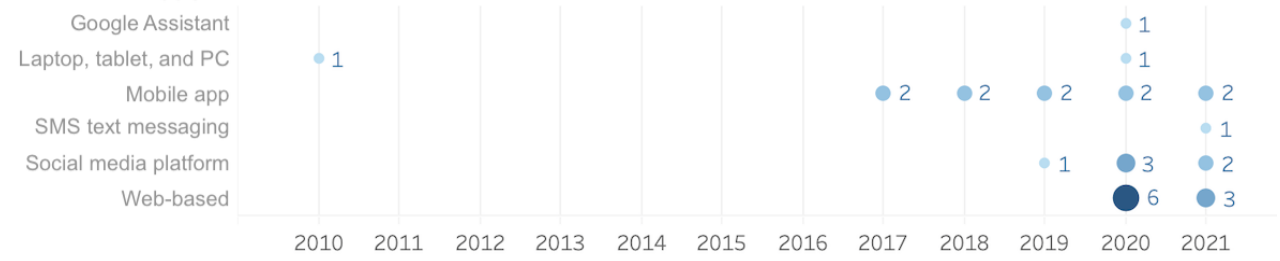


Figure 5. The ways in which users could message the chatbot were either by choosing from a set of predefined options or freely typing text as in a typical messaging app.

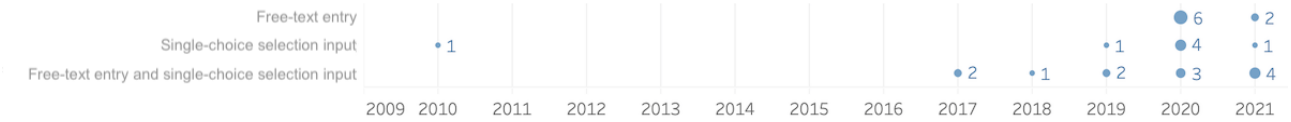
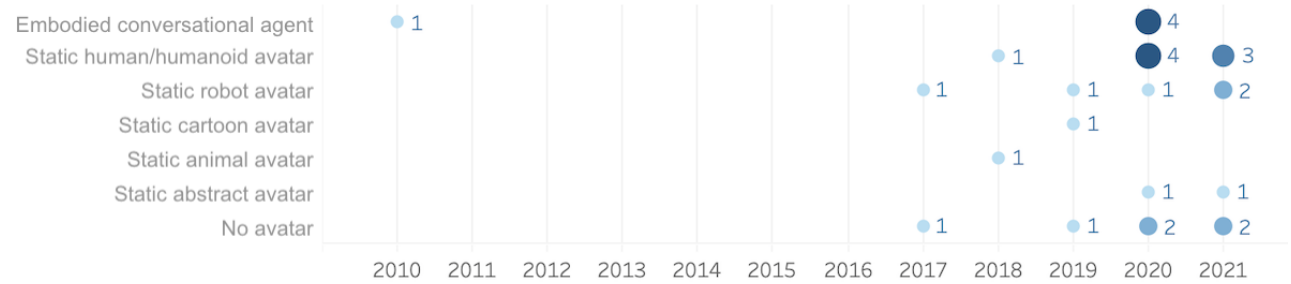


Figure 6. Presentation of the chatbot avatar.



Evidence for the Efficacy of Chatbot-Based Health Interventions

The included studies consisted of RCTs (n=12), user analytics (n=8), user experience studies (n=3), an experimental pilot (n=1), a descriptive study (n=1), comparative analyses (n=2), a case control study (n=1), design processes (n=2), and feasibility studies (n=2).

For RCTs, the number of participants varied between 20 to 927, whereas user analytics studies considered data from between 129 and 36,070 users. Overall, the evidence found was positive, showing some beneficial effect, or mixed, showing little or no

effect. Evidence was predominantly preliminary and context specific. Most (21/32, 65%) of the included studies established that the chatbots were usable but with some differences in the user experience and that they can provide some positive support across the different health domains.

Moderate positive results were found across several studies in regard to knowledge and skill improvement [20,39], reducing health risks [25], and supporting diet and physical exercise [31], and there is some preliminary evidence of chatbots that support smoking cessation improving the chances of quitting [36].

Studies on the use of chatbots for mental health, in particular anxiety and depression, also seem to show potential, with users reporting positive outcomes on at least some of the measurements taken [33,34,41]. The research suggests that psychotherapy chatbots can act as a supplemental tool as part of the broader psychotherapy process [21] across a broad range of psychotherapeutic methodologies and approaches (see Multimedia Appendix 2 for a summary of chatbot roles).

Chatbots were found to have improved medical service provision by reducing screening times [17] and triaging people with COVID-19 symptoms to direct them toward testing if required. These studies clearly indicate that chatbots were an effective tool for coping with the large numbers of people in the early stages of the COVID-19 pandemic. However, 1 comparative study [22] showed that the number of correctly assessed cases of COVID-19 varied considerably between the 10 web-based symptom checkers, with only 2 chatbots having a good balance between sensitivity (not classifying almost all patients as COVID-19–positive) and specificity (not classifying all patients as COVID-19–negative). Overall, this result suggests that although chatbots can achieve useful scalability properties (handling many cases), accuracy is of active concern, and their deployment needs to be evidence-based [23].

The evidence for the use of chatbots to support behavior change is mixed. One study found that any effect was limited to users who were already contemplating such change [24], and another study provided preliminary evidence for a health coach in older adults [31]. Another study reported finding no significant effect on supporting problem gamblers despite high completion rates [40].

Mixed findings were also reported regarding adherence. One study found that there was no effect on adherence to a blood pressure–monitoring schedule [39], whereas another reported a positive improvement medication adherence [35].

Research on the use of chatbots in public health service provision is at an early stage. Although preliminary results do indicate positive effects in a number of application domains, reported findings are for the most part mixed. Moreover, varying user engagement with the chatbots (though not necessarily correlated with the effect [36]), the size of the study, and the demographic characteristics of the target population (eg, some groups of people were more likely to have a better experience using the chatbot [18]) are some of the few variables that might affect the efficacy of an intervention.

Evaluation of Chatbot Design

The majority (26/32, 81%) of the studies used quantitative methods and methodologies for the evaluation of chatbot design and their impact in relation to health outcomes. For the most part, qualitative methods were used to examine the acceptability of chatbots to patients and their self-reported experience in using them alongside other quantitative usability metrics [45]. User experience and usability evaluation consisted of structured questionnaires and surveys, usually with a few open-ended questions (n=11), and just 1 study used a focus group (n=1).

By far the most prevalent means of assessing health impacts of chatbot-led interventions were RCTs (n=12). Studies that focused on the effectiveness of chatbots with regard to an assigned task, such as triage and symptom checking, lent themselves more easily to evaluation through user analytics (n=8). There was, however, limited evaluation of user experience (n=3), and chatbot development was rarely design-led—although there were notable exceptions, with 1 study identifying user principles in development [17]; 1 study following a human-centered design process with young adults treated for cancer [43]; and another using community based, participatory research to develop a chatbot [16]. A further limitation noted in the evaluation of widely deployed chatbots is that the data collected by the chatbot for further analysis do not hold personally identifiable information, so it is not possible to know if the targeted population are the actual users [16].

No included studies reported direct observation (in the laboratory or in situ; eg, ethnography) or in-depth interviews as evaluation methods. Given the recognized need for observational study in chatbots deployed for public health [28] and the current widespread use of observational and participatory methodologies in human-computer interaction (HCI) [47], there is an impetus for future chatbot research to rely on such methodologies if their development is to best support their users.

Discussion

Principal Findings

Although research on the use of chatbots in public health is at an early stage, developments in technology and the exigencies of combatting COVID-19 have contributed to the huge upswing in their use, most notably in triage roles. Studies on the use of chatbots for mental health, in particular depression, also seem to show potential, with users reporting positive outcomes [33,34,41]. Impetus for the research on the therapeutic use of chatbots in mental health, while still predominantly experimental, predates the COVID-19 pandemic. However, the field of chatbot research is in its infancy, and the evidence for the efficacy of chatbots for prevention and intervention across all domains is at present limited.

Notably, people seem more likely to share sensitive information in conversation with chatbots than with another person [20]. Speaking with a chatbot and not a person is perceived in some cases to be a positive experience as chatbots are seen to be less “judgmental” [48]. Human-like interaction with chatbots seems to have a positive contribution to supporting health and

well-being [27] and countering the effects of social exclusion through the provision of companionship and support [49]. However, in other domains of use, concerns over the accuracy of AI symptom checkers [22] framed the relationships with chatbot interfaces. The trustworthiness and accuracy of information were factors in people abandoning consultations with diagnostic chatbots [28], and there is a recognized need for clinical supervision of the AI algorithms [9].

Although the COVID-19 pandemic has driven the use of chatbots in public health, of concern is the degree to which governments have accessed information under the rubric of security in the fight against the disease. For example, in South Korea, the implementation of integrated technological responses, including personalized communication chatbots and the use of personal data gathered for contact tracing [50], uses AI in a way that transgresses what many would argue are fundamental human rights to privacy. The sharing of health data gathered through symptom checking for COVID-19 by commercial entities and government agencies presents a further challenge for data privacy laws and jurisdictional boundaries [51].

The evidence cited in most of the included studies either measured the effect of the intervention or surface and self-reported user satisfaction. There was little qualitative experimental evidence that would offer more substantive understanding of human-chatbot interactions, such as from participant observations or in-depth interviews. In this respect, we should remember that chatbots are complex systems, and chatbot deployment in public health is a technology design activity (the design of the platform, the communication modality, and content), as much as it is a medical intervention (the design of the intervention and setting up measures for its effectiveness). As an interdisciplinary subject of study for both HCI and public health research, studies must meet the standards of both fields, which are at times contradictory [52]. Methods developed for the evaluation of pharmacological interventions such as RCTs, which were designed to assess the effectiveness of an intervention, are known in HCI and related fields [53] to be limited in the insights they provide toward better design.

Studies in the existing research often do not provide sufficient information about the design of the chatbot being tested to be reproducible, including by RCT standards, as the chatbot description is not sufficient for an equivalent chatbot to be implemented. There are further confounding factors in the intervention design that are not directly chatbot related (eg, daily notifications for inputting mood data) or include aspects such as the chatbot's programmed personality that affect people differently [33]. As an emerging field of research, the future implications of human interactions with AI and chatbot interfaces is unpredictable, and there is a need for standardized reporting, study design [54,55], and evaluation [56].

Few of the included studies discussed how they handled safeguarding issues, even if only at the design stage. Of those that did, the studies mentioned that they could not provide a person to support the chatbot (ie, conversations with the chatbot are not monitored by a person), so the chatbot was programmed to message the user to contact official health authorities if they had an issue (eg, directing the user to call 911). This methodology is a particular concern when chatbots are used at scale or in sensitive situations such as mental health. In this respect, chatbots may be best suited as supplements to be used alongside existing medical practice rather than as replacements [21,33].

Implications for Future Research

Although the use of NLP is a new territory in the health domain [47], it is a well-studied area in computer science and HCI. When developing and deploying new technological interventions, one must take care to identify the ways in which these interventions might replicate or amplify existing inequities, such as access to language proficiency, technology literacy, smartphone technology, mobile data, and even electricity [9]. Human-centered design processes used in HCI and computer science, particularly those that engage the target user throughout the design process such as participatory design, co-design, and participatory action research, could be useful methods for addressing existing inequities from the beginning [57].

Most of the included papers contained screenshots of the chatbots. However, some of these were sketches of the interface rather than the final user interface, and most of the screenshots had insufficient description as to what the capabilities were. Although the technical descriptions of chatbots might constitute separate papers in their own right, these descriptions were outside the scope for our focus on evidence in public health. However, a previously published scoping review [58], focusing on the technical aspects of chatbots' implementation for medical use, distinguished between text-understanding modality (eg, pattern matching, machine learning, fixed input, and hybrid), data management (medical knowledge database, user information database, and conversation scripts), and text generation (fixed output and machine learning). A further scoping study would be useful in updating the distribution of the technical strategies being used for COVID-19–related chatbots.

Future research on chatbots would benefit from including more details as to how the chatbot is implemented and what type of NLP it uses and cross-referencing the equivalent technical paper describing the system implementation and technical contribution, if it is available.

More broadly, in a rapidly developing technological field in which there is substantial investment from industry actors, there is a need for better reporting frameworks detailing the technologies and methods used for chatbot development. Similarly, given the huge range of chatbot deployments across a wide variety of public health domains, there is a need for standards of comparative criteria to facilitate a better evaluation and validation of these agents and the methods and approaches that they use to improving health and well-being. Finally, there is a need to understand and anticipate the ways in which these technologies might go wrong and ensure that adequate safeguarding frameworks are in place to protect and give voice to the users of these technologies.

Limitations

Given the immaturity of the research on chatbots, the huge investment in their development and use for health, and the

dynamic nature of AI and HCI, our study does not capture the abundance of chatbots, commercial and otherwise, that have been developed across of the domains of public health application. There is a substantial lag between the production of academic knowledge on chatbot design and health impacts and the progression of the field.

Conclusions

Research on the recent advances in AI that have allowed conversational agents more realistic interactions with humans is still in its infancy in the public health domain. Studies show potential, especially for easily automated and repetitive tasks, but at the same time, concerns with the clinical, legal, and ethical aspects of the use of conversational agents for health care are well founded given the speed with which they have been adopted in practice. There is still little evidence in the form of clinical trials and in-depth qualitative studies to support widespread chatbot use, which are particularly necessary in domains as sensitive as mental health. Most of the chatbots used in supporting areas such as counseling and therapeutic services are still experimental or in trial as pilots and prototypes. Where there is evidence, it is usually mixed or promising, but there is substantial variability in the effectiveness of the chatbots. This finding may in part be due to the large variability in chatbot design (such as differences in content, features, and appearance) but also the large variability in the users' response to engaging with a chatbot.

There is no doubting the extent to which the use of AI, including chatbots, will continue to grow in public health. The ethical dilemmas this growth presents are considerable, and we would do well to be wary of the enchantment of new technologies [59]. Of paramount concern is the need to understand where we can use automation over other technologies that connect humans to humans (eg, machine assistance instead of machine intelligence) and what are the situations in which a conversation with a computer that simulates another person is indeed serving the needs of the person. For example, the recently published WHO Guidance on the Ethics and Governance of AI in Health [10] is a big step toward achieving these goals and developing a human rights framework around the use of AI. However, as Privacy International commented in a review of the WHO guidelines, the guidelines do not go far enough in challenging the assumption that the use of AI will inherently lead to better outcomes [60]. Digital innovation in public health should ideally be informed by research that measures the impact that technologies such as chatbots may have in health interventions, provides insight into user experience, and works to ensure the safety of and promote the well-being of users.

Conflicts of Interest

The original study that is the basis of this paper was commissioned by the World Health Organization. MM is a contractor of Katikati, a social technology start-up enabling 1-to-1 human-led conversations at scale over SMS text messaging, instant messaging, or web.

Multimedia Appendix 1
Search terms.
[DOCX File , 15 KB - humanfactors_v9i4e35882_app1.docx]

Multimedia Appendix 2
Chatbot application domain, purpose, interaction type, and findings summary.
[DOCX File , 33 KB - humanfactors_v9i4e35882_app2.docx]

References

1. Rajan A. Amol Rajan interviews Sundar Pichai, in Amol Rajan interviews. BBC Radio 2. URL: https://www.bbc.co.uk/iplayer/episode/m000xwll/amol-rajan-interviews-sundar-pichai [accessed 2021-07-22]
2. Adamopoulou E, Moussiades L. An overview of chatbot technology. In: IFIP Advances in Information and Communication Technology, vol 584. 2020 May 29 Presented at: AIAI 2020: Artificial Intelligence Applications and Innovations; June 5-7, 2020; Neos Marmaras, Greece p. 373-383. [doi: 10.1007/978-3-030-49186-4_31]
3. Brandtzaeg PB, Følstad A. Chatbots: changing user needs and motivations. Interactions 2018 Aug 22;25(5):38-43. [doi: 10.1145/3236669]
4. Borah B, Pathak D, Sarmah P, Som B, Nandi S. Survey of textbased chatbot in perspective of recent technologies. In: Communications in Computer and Information Science, vol 1031. 2019 Jun 26 Presented at: CICBA 2018: Computational Intelligence, Communications, and Business Analytics; July 27-28, 2018; Kalyani, India p. 84-96. [doi: 10.1007/978-981-13-8581-0_7]
5. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
6. McKillop M, South BR, Preininger A, Mason M, Jackson GP. Leveraging conversational technology to answer common COVID-19 questions. J Am Med Inform Assoc 2021 Mar 18;28(4):850-855 [FREE Full text] [doi: 10.1093/jamia/ocaa316] [Medline: 33517402]

7. Almalki M, Azeez F. Health chatbots for fighting COVID-19: a scoping review. Acta Inform Med 2020 Dec;28(4):241-247 [FREE Full text] [doi: 10.5455/aim.2020.28.241-247] [Medline: 33627924]
8. Höhn S, Bongard-Blanchy K. Heuristic evaluation of COVID-19 chatbots. In: Lecture Notes in Computer Science, vol 12604. 2021 Feb 03 Presented at: CONVERSATIONS 2020: Chatbot Research and Design; November 23-24, 2020; Virtual event p. 131-144. [doi: 10.1007/978-3-030-68288-0_9]
9. Lai L, Wittbold KA, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: novel strategies for population health management in response to the COVID-19 pandemic. Healthc (Amst) 2020 Dec;8(4):100493 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100493] [Medline: 33129176]
10. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. 2021 Jun 28. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2022-08-30]
11. Walwema J. The WHO health alert: communicating a global pandemic with WhatsApp. J Bus Tech Commun 2020 Sep 16;35(1):35-40. [doi: 10.1177/1050651920958507]
12. WHO launches a chatbot on Facebook Messenger to combat COVID-19 misinformation. World Health Organization. 2020 Apr 15. URL: https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation [accessed 2021-07-21]
13. Gabarron E, Larbi D, Denecke K, Årsand E. What do we know about the use of chatbots for public health? Stud Health Technol Inform 2020 Jun 16;270:796-800. [doi: 10.3233/SHTI200270] [Medline: 32570492]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Koplan JP, Bond TC, Merson MH, Reddy KS, Rodriguez MH, Sewankambo NK, Consortium of Universities for Global Health Executive Board. Towards a common definition of global health. Lancet 2009 Jun 06;373(9679):1993-1995. [doi: 10.1016/S0140-6736(09)60332-9] [Medline: 19493564]
16. Bonnevie E, Lloyd T, Rosenberg S, Williams K, Goldbarg J, Smyser J. Layla's Got You: developing a tailored contraception chatbot for Black and Hispanic young women. Health Educ J 2020 Dec 18;80(4):413-424. [doi: 10.1177/0017896920981122]
17. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
18. Roman M, Bellei EA, Biduski D, Pasqualotti A, De Araujo CDSR, De Marchi ACB. "Hey assistant, how can I become a donor?" the case of a conversational agent designed to engage people in blood donation. J Biomed Inform 2020 Jul;107:103461 [FREE Full text] [doi: 10.1016/j.jbi.2020.103461] [Medline: 32504669]
19. Bechmann A, Bowker G. Unsupervised by any other name: hidden layers of knowledge production in artificial intelligence on social media. Big Data & Society 2019 Feb 05;6(1):205395171881956. [doi: 10.1177/2053951718819569]
20. Maeda E, Miyata A, Boivin J, Nomura K, Kumazawa Y, Shirasawa H, et al. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. Reprod Biomed Online 2020 Dec;41(6):1133-1143. [doi: 10.1016/j.rbmo.2020.09.006] [Medline: 33039321]
21. Xu B, Zhuang Z. Survey on psychotherapy chatbots. Concurr Comput 2020 Dec 28;34(7):2020. [doi: 10.1002/cpe.6170]
22. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. J Med Internet Res 2020 Oct 06;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
23. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
24. Gardiner P, Bickmore T, Yinusa-Nyahkoon L, Reichert M, Julce C, Sidduri N, et al. Using health information technology to engage African American women on nutrition and supplement use during the preconception period. Front Endocrinol (Lausanne) 2020 Jan 19;11:571705 [FREE Full text] [doi: 10.3389/fendo.2020.571705] [Medline: 33584534]
25. Jack BW, Bickmore T, Yinusa-Nyahkoon L, Reichert M, Julce C, Sidduri N, et al. Improving the health of young African American women in the preconception period using health information technology: a randomised controlled trial. Lancet Digit Health 2020 Sep;2(9):e475-e485 [FREE Full text] [doi: 10.1016/S2589-7500(20)30189-8] [Medline: 33328115]
26. Hautz WE, Exadaktylos A, Sauter TC. Online forward triage during the COVID-19 outbreak. Emerg Med J 2021 Mar 11;38(2):106-108 [FREE Full text] [doi: 10.1136/emermed-2020-209792] [Medline: 33310732]
27. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020 Mar 06;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
28. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of self-diagnosis health chatbots in real-world settings: case study. J Med Internet Res 2021 Jan 06;23(1):e19928 [FREE Full text] [doi: 10.2196/19928] [Medline: 33404508]
29. Oh YH, Chung K, Ju DY. Differences in interactions with a conversational agent. Int J Environ Res Public Health 2020 May 04;17(9):3189 [FREE Full text] [doi: 10.3390/ijerph17093189] [Medline: 32375330]
30. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]

31. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
32. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. JMIR Form Res 2021 May 11;5(5):e27868 [FREE Full text] [doi: 10.2196/27868] [Medline: 33973854]
33. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M, et al. Pocket Skills: a conversational mobile web app to support dialectical behavioral therapy. 2018 Apr 21 Presented at: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC p. 1-15. [doi: 10.1145/3173574.3173972]
34. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth uHealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
35. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
36. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? an experimental study. Digit Health 2019 Sep 30;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
37. Bickmore T, Ring L. Making it personal: end-user authoring of health narratives delivered by virtual agents. In: Lecture Notes in Computer Science, vol 6356. 2010 Presented at: IVA 2010: Intelligent Virtual Agents; September 20-22, 2010; Philadelphia, PA p. 399-405. [doi: 10.1007/978-3-642-15892-6_43]
38. Bickmore T, Rubin A, Simon S. Substance use screening using virtual agents: towards automated screening, brief intervention, and referral to treatment (SBIRT). 2020 Oct 19 Presented at: IVA '20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents; October 20-22, 2020; Virtual event, Scotland, UK p. 1-7. [doi: 10.1145/3383652.3423869]
39. Echeazarra L, Pereira J, Saracho R. TensioBot: a chatbot assistant for self-managed in-house blood pressure checking. J Med Syst 2021 Mar 15;45(4):54. [doi: 10.1007/s10916-021-01730-x] [Medline: 33723721]
40. So R, Furukawa TA, Matsushita S, Baba T, Matsuzaki T, Furuno S, et al. Unguided chatbot-delivered cognitive behavioural intervention for problem gamblers through messaging app: a randomised controlled trial. J Gambl Stud 2020 Dec 11;36(4):1391-1407. [doi: 10.1007/s10899-020-09935-4] [Medline: 32162075]
41. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial intelligence-based chatbot for anxiety and depression in university students: pilot randomized controlled trial. JMIR Form Res 2021 Aug 12;5(8):e20678 [FREE Full text] [doi: 10.2196/20678] [Medline: 34092548]
42. Dosovitsky G, Pineda BS, Jacobson NC, Chang C, Escoredo M, Bunge EL. Artificial intelligence chatbot for depression: descriptive study of usage. JMIR Form Res 2020 Nov 13;4(11):e17065 [FREE Full text] [doi: 10.2196/17065] [Medline: 33185563]
43. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR mHealth uHealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
44. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR mHealth uHealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
45. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
46. Espinoza J, Crown K, Kulkarni O. A guide to chatbots for COVID-19 screening at pediatric health care facilities. JMIR Public Health Surveill 2020 Apr 30;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]
47. Lazar J, Feng J, Hochheiser H. Research Methods in Human-Computer Interaction. 2nd ed. Burlington, MA: Morgan Kaufmann; Apr 19, 2017.
48. Jack B, Bickmore T, Hempstead M, Yinusa-Nyahkoon L, Sadikova E, Mitchell S, et al. Reducing preconception risks among African American women with conversational agent technology. J Am Board Fam Med 2015 Jul 07;28(4):441-451 [FREE Full text] [doi: 10.3122/jabfm.2015.04.140327] [Medline: 26152434]
49. de Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. Front Psychol 2019 Jan 23;10:3061 [FREE Full text] [doi: 10.3389/fpsyg.2019.03061] [Medline: 32038415]
50. Nam T. How did Korea use technologies to manage the COVID-19 crisis? a country report. International Review of Public Administration 2020 Nov 20;25(4):225-242. [doi: 10.1080/12294659.2020.1848061]
51. Chouhan A, Pathak S, Tendulkar R. Chatbots for coronavirus: detecting COVID-19 symptoms with virtual assessment tool. In: Kautish S, Peng SL, Obaid AJ, editors. Computational Intelligence Techniques for Combating COVID-19. Cham, Switzerland: Springer; May 01, 2021:275-304.

52. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018 May 03;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
53. Lewis L. Avatars and robots as social companions in healthcare: requirements, engineering, adoption and ethics. In: Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications. Hershey, PA: IGI Global; 2017:582-602.
54. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
55. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
56. Kearns WR, Chi N, Choi YK, Lin S, Thompson H, Demiris G. A systematic review of health dialog systems. Methods Inf Med 2019 Dec 29;58(6):179-193. [doi: 10.1055/s-0040-1708807] [Medline: 32349152]
57. Neuhauser L. Integrating participatory design and health literacy to improve research and interventions. Stud Health Technol Inform 2017;240:303-329. [doi: 10.3233/978-1-61499-790-0-303] [Medline: 28972525]
58. Safi Z, Abd-Alrazaq A, Khalifa M, Househ M. Technical aspects of developing chatbots for medical applications: scoping review. J Med Internet Res 2020 Dec 18;22(12):e19127 [FREE Full text] [doi: 10.2196/19127] [Medline: 33337337]
59. Madianou M. Nonhuman humanitarianism: when 'AI for good' can be harmful. Information, Communication & Society 2021 Apr 08;24(6):850-868. [doi: 10.1080/1369118X.2021.1909100]
60. Our analysis of the WHO report on Ethics and Governance of Artificial Intelligence for Health. Privacy International. 2021 Jul 20. URL: https://privacyinternational.org/news-analysis/4594/our-analysis-who-report-ethics-and-governance-artificial-intelligence-health [accessed 2021-07-30]

Abbreviations

AI: artificial intelligence
HCI: human-computer interaction
NLP: natural language processing
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
WHO: World Health Organization

Edited by A Kushniruk; submitted 21.12.21; peer-reviewed by H Tanaka, J Parsons; comments to author 23.03.22; revised version received 14.07.22; accepted 02.08.22; published 05.10.22.

Please cite as:
Wilson L, Marasoiu M
The Development and Use of Chatbots in Public Health: Scoping Review
JMIR Hum Factors 2022;9(4):e35882
URL: https://humanfactors.jmir.org/2022/4/e35882
doi:10.2196/35882
PMID:36197708

Review

Acceptability and Feasibility of Wearable Transdermal Alcohol Sensors: Systematic Review

Eileen Brobbin[1], BSc, MSc; Paolo Deluca[1], PhD; Sofia Hemrage[1], BSc, MSc; Colin Drummond[1], MB, ChB

Department of Addictions, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Eileen Brobbin, BSc, MSc
Department of Addictions
Institute of Psychiatry, Psychology & Neuroscience
King's College London
Addiction Science Building
4 Windsor Walk
London, SE5 8BB
United Kingdom
Phone: 44 0207 836 545
Email: eileen.brobbin@kcl.ac.uk

Abstract

Background: Transdermal alcohol sensors (TASs) have the potential to be used to monitor alcohol consumption objectively and continuously. These devices can provide real-time feedback to the user, researcher, or health professional and measure alcohol consumption and peaks of use, thereby addressing some of the limitations of the current methods, including breathalyzers and self-reports.

Objective: This systematic review aims to evaluate the acceptability and feasibility of the currently available TAS devices.

Methods: A systematic search was conducted in CINAHL, EMBASE, Google Scholar, MEDLINE, PsycINFO, PubMed, and Scopus bibliographic databases in February 2021. Two members of our study team independently screened studies for inclusion, extracted data, and assessed the risk of bias. The study's methodological quality was appraised using the Mixed Methods Appraisal Tool. The primary outcome was TAS acceptability. The secondary outcome was feasibility. The data are presented as a narrative synthesis.

Results: We identified and analyzed 22 studies. Study designs included laboratory- and ambulatory-based studies, mixed designs, randomized controlled trials, and focus groups, and the length the device was worn ranged from days to weeks. Although views on TASs were generally positive with high compliance, some factors were indicated as potential barriers and there are suggestions to overcome these.

Conclusions: There is a lack of research investigating the acceptability and feasibility of TAS devices as a tool to monitor alcohol consumption in clinical and nonclinical populations. Although preliminary evidence suggests their potential in short-term laboratory-based studies with volunteers, more research is needed to establish long-term daily use with other populations, specifically, in the clinical and the criminal justice system.

Trial Registration: PROSPERO CRD42021231027; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=231027

(JMIR Hum Factors 2022;9(4):e40210) doi:10.2196/40210

KEYWORDS

alcohol consumption; alcohol monitoring; digital technology; transdermal alcohol sensors; wearables; acceptability; feasibility; monitoring; sensors; real-time feedback; health promotion; alcohol intake

Introduction

Background

In recent years, there have been advances in the development of various wearable transdermal alcohol sensor (TAS) devices. These devices measure alcohol consumption from vapors off the skin via sweat, known as transdermal alcohol concentration. The advantage of these devices is the possibility to wear them all day, allowing for long-term continuous data capture [1] and recording drinking episodes accurately in near real time. These are features that overcome some of the limitations of other

methods. Breath and urine analysis is limited by the short half-life of ethanol [2]. Blood markers of heavy alcohol consumption (eg, gamma-glutamyl transferase, liver function tests, mean cell volume, carbohydrate-deficient transferrin) have limitations in terms of sensitivity or specificity [2,3]. Although other alcohol biomarkers such as phosphatidyl ethanol found in blood and ethyl glucuronide and ethyl sulfate found in urine have been found to have high sensitivity and specificity, these tests can only be used to detect if alcohol has been consumed within the last number of days [4-6]. Further, all these tests require administration by a trained health professional. In contrast, TAS devices are noninvasive, objective, easy-to-use, low-maintenance, and allow for the study of behaviors in real-world contexts for potentially weeks or months at a time [7].

The newer TAS devices have the potential to communicate wirelessly through a sim card or with a smartphone [7]. Using this as a method of data collection could reduce the time and resources required for data capture. There is also the possibility for information to be uploaded over a mobile network and delivered to the patient, health professional, or researcher in near real time. Evidence suggests there is a slight time lag between drinking and peak transdermal alcohol concentration [7-10], but preliminary studies with newer generation devices have demonstrated a reduction in this time lag [7,9]. SCRAM is a chunkier device with various types worn on the ankle and is similar in appearance to a house arrest monitor, whereas WrisTAS, BACtrack, ION, and Quantac Tally are worn exclusively on the wrist and are smaller in style, closer in appearance to a FitBit, health watch, or pedometer, and are approximately the same size as a watch (information along with an image of each device is summarized in Multimedia Appendix 1).

Potential Uses of TAS Devices

Alcohol Treatment and Interventions

TAS devices have the potential to improve clients' engagement, increase clinicians' ability to accurately assess consumption, and trigger real-time interventions. If relapses are caught early, the treatment service could personalize treatment and interventions for the client to prevent further or larger relapses [11]. In addition, TASs can capture regular data, which can be linked with contingency management (CM) for an effective treatment option [12-14].

Forensic Monitoring

The South Dakota 24/7 Sobriety initiative enforced alcohol monitoring for driving under the influence offenders. The use of TASs, breathalyzers, and sanctions for breaches was found to be effective. This project included 17,000 individuals between 2005 and 2010, and since then, has been extended to domestic violence and drug offences [15]. There has been some preliminary research implementing remote alcohol monitoring within the United Kingdom [16]. Most recently, England and Wales announced new legislation, where alcohol-related offenders may be banned from drinking alcohol and be ordered to wear a TAS for up to 120 days.

Research

Alcohol research mostly relies on retrospective self-report data. There is some evidence of the reliability of self-report [17]; however, it can be subject to bias [18,19]. Evidence suggests that alcohol consumption tends to be underreported [20] and may be more greatly underreported with nonroutine drinking patterns and heavy drinkers [21], whereas TASs could provide an objective measure.

Public Use

Newer TAS devices are designed for consumer use. They could be used to monitor alcohol levels before driving, as proof of sobriety at bars or public events, and be a diary for those interested in monitoring alcohol consumption for general health.

Acceptability and Feasibility of TAS Devices

The acceptability and feasibility of health care interventions are important issues to consider in their development, evaluation, and implementation [22]. Although previous studies [7,12,23-25] have alluded to the acceptability and feasibility of TASs to objectively monitor alcohol consumption, there are a limited number of studies addressing this, and to our knowledge, there are no systematic reviews specifically investigating this. In this review, we consider acceptability as the device being perceived as appropriate, which is based on both cognitive and emotional responses to the devices and that this acceptability can be assessed before, during, or after wearing the device [22]. We consider feasibility as the extent to which this device could be implemented practically within the identified setting. This systematic review investigates the current knowledge by systematically identifying and evaluating the existing literature on the use of TAS devices in clinical and nonclinical populations, alone or in conjunction with a psychosocial intervention. The objective of this review was to assess the acceptability and feasibility of TAS devices with an overarching objective of establishing the barriers and facilitators to implementing these devices.

Methods

Systematic Review Registration

This systematic review has been written according to the PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis for Protocols) guidelines [26]. This protocol has been registered on the Prospective Register of Systematic Reviews (PROSPERO CRD42021231027). On review of the results, it was decided that the findings of this systematic review should be reported in 2 papers: this paper focusing on acceptability and feasibility outcomes and another paper on accuracy outcomes [27].

Inclusion Criteria for the Studies

Studies meeting all the following criteria were included: full text, original studies, published in peer-reviewed journals, written in English, using a wearable transdermal sensor device(s) or investigating attitudes and experience of TAS use reporting acceptability or feasibility outcomes. There were no restrictions on publication year or participant clinical characteristics. Data

based on conference abstracts, dissertations, and grey literature were not included.

Information Sources

Bibliographic databases included CINAHL, EMBASE, Google Scholar, MEDLINE, PsycINFO, PubMed, and Scopus. Searches were carried out February 2021 (Textbox 1 and Multimedia Appendix 2 and Multimedia Appendix 3). The searches were supplemented by cross-checking the reference lists of key publications, related systematic reviews, and all included papers. All identified titles and abstracts were screened in Covidence. From this list, the full text was retrieved and assessed for eligibility (EB). Any queries were discussed with a second reviewer (SH). Any disagreement was resolved through discussion with a third reviewer (PD). A data extraction form was created, piloted, and refined as necessary. EB extracted the data independently (Multimedia Appendix 4) and the second reviewer completed entries check for accuracy.

Textbox 1. Search terms.

- Transdermal alcohol sensor
- Transdermal alcohol sensor device
- Transdermal alcohol monitoring device
- Transdermal alcohol bracelet
- Transdermal alcohol wristband
- Transdermal alcohol ankle
- Transdermal alcohol concentration
- Transdermal alcohol concentration data
- Transdermal alcohol sensor data
- Transdermal alcohol validity
- Transdermal alcohol acceptability
- Transdermal alcohol feasibility

Outcomes in the Studies

All outcome measures reported in the included studies were extracted, including both objective and self-reported measures. The definition of acceptability outcomes for TASs in the context of this study are factors that affect participant willingness to use the device in treatment or rehabilitation. The definition of feasibility outcomes for TASs in the context of this study are factors that would impact the introduction, including the operational capability of these devices in alcohol treatment services as part of treatment and individual skills required by wearers.

Quality Assessment

We used the Mixed Methods Appraisal Tool (MMAT), as it was designed for the appraisal in reviews that include a range of designs (qualitative, quantitative, and mixed methods) [28]. For each included study, we determined the study design and then applied the appropriate screening criteria; this provided an overall quality score for the study [28]. EB independently completed the appraisal, and queries were discussed with a second reviewer (SH). Any disagreements were resolved by discussion with a third reviewer (PD).

Data Synthesis and Analyses

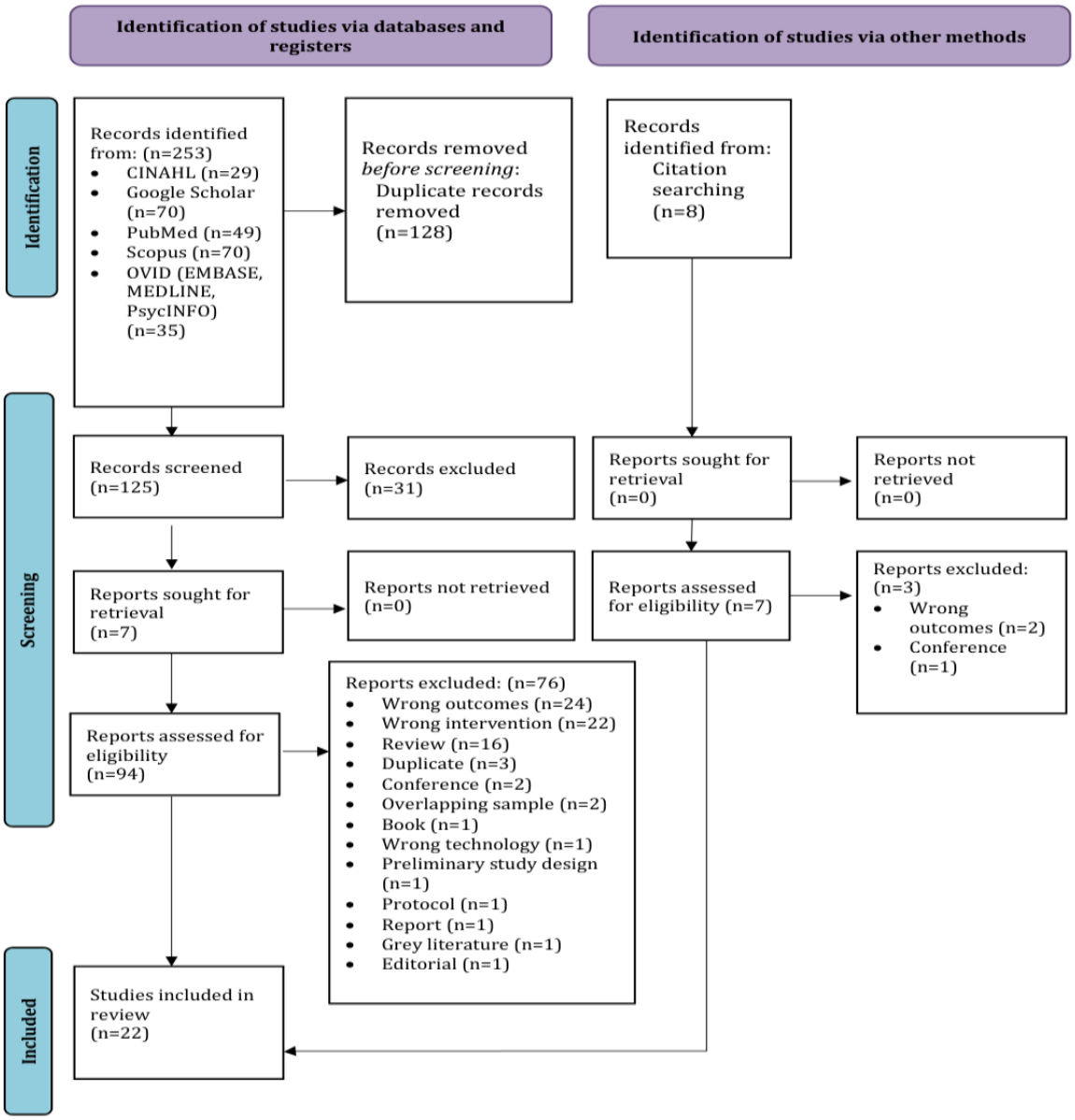
The data are summarized using a structured narrative description. As there are no standard outcomes for acceptability or feasibility measures for TASs, we include any acceptability and feasibility measures they report. We found that outcomes fell into one of the following categories and thus, these are used as subheadings in the results narrative: comfort, appearance, ease of use, social perceptions, perceptions of alcohol use, barriers/suggestions, the criminal justice system, device tampering, and compliance.

Results

Studies in This Review

After removing duplicates, 125 papers were screened, 31 papers were excluded at the title and abstract screening, and 94 full-text papers were assessed for eligibility. A total of 76 papers were then excluded. There were 8 additional papers identified by citation searching; of these, 5 were included. The final sample included 22 publications (Figure 1).

Figure 1. PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis for Protocols) flow diagram.



Characteristics of the Studies

Sixteen studies used a version of SCRAM [8,12,14,23-25,29-38], 4 used a version of WrisTAS [39-42], 3 used BACtrack Skyn [7,38,43], and 1 used Quantac Tally [7]. Some studies used more than 1 version or brand of device. Almost half of the included studies (10/22, 45%) aimed to assess how TASs can be used to measure alcohol consumption. Out of the 22 studies, 8 (36%) aimed to assess acceptability, adherence, and feasibility with TASs, and 5 (23%) used TASs to explore their effectiveness in implementing CM for alcohol reduction treatment and evaluating the efficacy of CM for alcohol use reduction. One investigated nonalcoholic energy drinks and one investigated alcohol-related positive mode enhancement and negative mood reduction.

Most studies (18/22, 82%) included participants who were adults in good health. Only 4 out of 22 (18%) studies included participants who were diagnosed with alcohol dependence. These 4 studies recruited participants from community-based clinics receiving alcohol treatment [8,23,24] or were recruited on admission to a general hospital substance abuse unit [42]. Most were conducted in the United States (17/22, 77%). The earliest paper included is from 1992 but the majority (17/22, 77%) of the studies were published since 2015.

There were 821 participants enrolled in total across the studies, with 793 included in the procedure or analysis. Therefore, 28 participants that were enrolled were not included in the results (28/821, 3.4%) (for reasons such as withdrawing or missing data). One paper [7] was still in the early stages of data collection for one of their studies and so these participant numbers were unknown. Not all studies included detailed information on the participants' age, gender, and ethnicity. Where information was provided, it could be seen that participants' ages ranged from 18 to 57 years, the majority included both females and males, and for most, Caucasian participants represented a large majority of the sample (see Table 1).

Table 1. Papers included in this review.

Author, year	Design	Aim	Population	Device; Participants: N=821 enrolled (N=793 included)	Age (years); Female (%); Caucasian (%)	Mixed Methods Appraisal Tool score (%)
Alessi et al [24], 2019	Ambulatory	Assess how we can measure alcohol consumption with this technology	Clinical: alcohol outpatient	SCRAMx; 66 (63)	Mean: 40.3; 41.3; 58.7	40
Alessi et al [23], 2017	Ambulatory	Assess acceptability, adherence, and feasibility with this technology	Clinical: alcohol outpatient	SCRAMx; 100 (98)	Mean: 42; 48; 59	80
Averill et al [25], 2018	Pilot RCT[a], Ambulatory	Effectiveness of TAM[b] in implementing CM[c] for alcohol reduction treatment in various population groups, evaluating the efficacy of CM reduction in alcohol use	Nonclinical: at risk drinkers	SCRAM; 37 (37)	Mean: 40.5; 0; 86.5	100
Ayala et al [29], 2009	Laboratory	Assessing nonalcoholic energy drinks with TAM	Nonclinical: good health	SCRAM-II; 15 (15)	Range: 22-35; 60; 80	80
Barnett et al [14], 2017	RCT, Ambulatory	Effectiveness of TAM in implementing CM for alcohol reduction treatment in various population groups, evaluating the efficacy of CM reduction in alcohol use	Nonclinical: heavy drinkers	SCRAM-II, SCRAMx; 30 (30)	Mean: 28.9, range: 21-57; 46; 76.7	80
Barnett et al [12], 2011	Ambulatory	Effectiveness of TAM in implementing CM for alcohol reduction treatment in various population groups, evaluating the efficacy of CM reduction in alcohol use	Nonclinical: heavy drinkers	SCRAM; 20 (13)	Mean: 32; 46; 69.2	80
Caluzzi et al [30], 2019	Ambulatory	Assess acceptability, adherence, and feasibility with this technology	Nonclinical: good health	SCRAM; 34 (30)	Range: 18-29; Sample 1: 58, sample 2: 50; Not known	100
Croff et al [39], 2020	Ambulatory	Assess acceptability, adherence, and feasibility with this technology	Nonclinical: good health	WrisTAS-7; 59 (57)	Mean: 18.82, range: 14-24; 100; 73.3	100
Fairbairn et al [31], 2018	Ambulatory	Alcohol-related positive mood enhancement and negative mood reduction study	Nonclinical: social drinkers	SCRAM; 48 (48)	Mean: 22.6, range: 21-28; 50; 56	100
Goodall et al [32], 2016	Focus groups	Assess acceptability, adherence, and feasibility with this technology	Serving offenders	SCRAM (RAM); 12 (12)	Not known; 0; Not known	100
Luczak et al [40], 2015	Mixed design	Assess how we can measure alcohol consumption with this technology	Nonclinical: good health	WrisTAS-7; 32 (32)	Mean: 23.1; 47; 0	80

Author, year	Design	Aim	Population	Device; Participants: N=821 enrolled (N=793 included)	Age (years); Female (%); Caucasian (%)	Mixed Methods Appraisal Tool score (%)
Mathias et al [33], 2018	Ambulatory	Effectiveness of TAM in implementing CM for alcohol reduction treatment in various population groups, evaluating the efficacy of CM reduction in alcohol use	Nonclinical: good health	SCRAM; 86 (86)	Mean: 38.5, 37.3, 39.3, 42.6 (cycle 1, 2, 3, 4); Cycle 1: 9, Cycle 2: 20, Cycle 3: 6, Cycle 4: 12; Not known	100
Neville et al [34], 2013	Ambulatory	Assess acceptability, adherence, and feasibility with this technology, assess how we can measure alcohol consumption with this technology	Nonclinical: good health	SCRAMx; 60 (53)	Mean: 21.46, range: 18-37; 0; Not known	80
Norman et al [35], 2020	Ambulatory	Assess acceptability, adherence, and feasibility with this technology, assess how we can measure alcohol consumption with this technology	Nonclinical: good health	SCRAM; 14 (14)	Mean: 21.9, range: 18-29; 50; Not known	80
Rash et al [36], 2019	Ambulatory	Assess how we can measure alcohol consumption with this technology	Nonclinical: heavy drinking	SCRAMx; 22 (19)	Mean: 46.8 (CM), 46 (monitoring); CM: 31, monitoring: 33; 23 (CM); 56 (monitoring)	100
Rosenberg et al [43], 2021	Ambulatory	Assess acceptability, adherence, and feasibility with this technology	Nonclinical: good health	BACtrack Skyn; 5 (5)	Mean: 21.6, range: 21.2-22.3; 60; Not known	80
Sakai et al [8], 2006	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	Alcohol dependent and nonalcohol-dependent	SCRAM; 44 (44)	Mean: 32.8, 38.1, 37.5 (lab: no, low, high dose), mean: 43.5, 39.9 (community: NAD[d], AD[e]); Laboratory study no/low/high dose: 50; community study NAD: 70, AD: 60; Not known	100
Simons et al [41], 2015	Ambulatory	Assess how we can measure alcohol consumption with this technology	Nonclinical: good health	WrisTAS-7; 60 (60)	Mean: 19.57, range: 18-21; 52; 97	80
Swift et al [42], 1992	Laboratory	Assess how we can measure alcohol consumption with this technology	Nonclinical: good health and alcohol dependent	WrisTAS; 15 (15)	Mean: 27, range: 21-40 (controlled); Range: 31-53 (intoxicated); Controlled: 20; intoxicated: 40; Not known	80

Author, year	Design	Aim	Population	Device; Participants: N=821 enrolled (N=793 included)	Age (years); Female (%); Caucasian (%)	Mixed Methods Appraisal Tool score (%)
Villalba et al [37], 2020	Focus groups	Assess acceptability, adherence, and feasibility with this technology, effectiveness of TAM in implementing CM for alcohol reduction treatment in various population groups, evaluating the efficacy of CM reduction in alcohol use	HIV-related community	SCRAM; 37 (37)	Not known; 51; Not known	100
Wang et al [7], 2019	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	Nonclinical: good health	Quantac Tally, BACtrack Skyn; Still recruiting	Not known; Not known; Not known	20
Wang et al [38], 2021	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	Nonclinical: good health	BACtrack Skyn, SCRAM-CAM; 25 (15)	Range: 36-38 (study 1); Mean 29.5 (study 2); Study 1: 33.3; study 2: 60; Not known	80

[a]RCT: randomized controlled trial.

[b]TAM: transdermal alcohol monitoring.

[c]CM: contingency management.

[d]NAD: non-alcohol dependent.

[2]AD: alcohol dependent

We found that no study defined acceptability and feasibility in terms of their research. One described how they measured protocol feasibility [40] and one described how they measured acceptability and feasibility of the device [43]. Luczak et al [40] report protocol feasibility as the reliability of each component of the protocol, the validity of TAS data, participant compliance, and reactivity. Rosenberg et al [43] report measuring acceptability using the Acceptability of Intervention Measure scale and feasibility with the amount of alcohol monitor data produced and the correlation between device-reported drinking events and drinking events reported by participants. They also used the Feasibility of Intervention Measure scale.

Quality Assessment

All studies, except for 2 studies [7,24], met a minimum of 4 out 5 MMAT criteria (>80%) (all scores reported in Table 1 and Multimedia Appendix 5). This is due to Alessi et al [24] not providing details about randomization and participant information. Alessi et al [24] included alcohol treatment outpatients who were randomized to usual care in 2 previous studies; these 2 previous studies are not clearly stated. Wang et al's study [7] was difficult to score due to incomplete data collection, as their study was still ongoing at the time of publication. With the MMAT, exclusion of low methodological quality studies is discouraged [28]. Due to the nature of many studies, blinding of participants or staff was not possible; in some, there were clear differences between groups [34] or CM incentives were provided [25], where the staff were required to know participant allocation. In many studies, there was only 1 group of participants who all completed the same task and so randomization was not required [7,8,12,23,29-33,35-43]. Not all studies provided clear information on participant information, randomization, incomplete outcome data, and selective reporting; therefore, there was potential bias due to limited information [7,12,14,23,24,29,34,35,38,40-43].

Acceptability Measures

Comfort

SCRAM is the device that is the biggest of the various models and typically worn on the ankle. Participants described rarely noticing SCRAM [8,23,25,30,36]; however, there were a few activities when the device was more noticeable, such as bathing, sport, sleep, and the device vibration/size impacting work and clothing choice [12,14,23,25,30,37]. When rating the SCRAM comfort, on average, it was rated as comfortable by wearers [25,36], and in another study, over half of the participants reported adjusting to any initial discomfort [30]. Other studies reported marks on the skin and itching caused by SCRAM [12,23,25]. Other physical side effects were reported as mild to moderate [14,23,30]. Rash et al [36] found reports of skin marks very uncommon with SCRAM; however, in the study of Caluzzi et al [30] participants described the clamp mechanism of SCRAM as constricting. Methods employed to increase comfort were adjustments made by researchers, including tube socks, plasters, medical tape, and supportive shoe wear. Participants felt that heat and dehydration increased discomfort and those with slimmer legs had greater difficulty keeping the device in a comfortable position. For WrisTAS, device marks were reported infrequently [40], and in 1 study, no marks were reported at all [42]. WrisTAS is smaller than SCRAM and is worn on the wrist. One study allowed participants to remove the device (WrisTAS-7) before engaging in activities that were

incompatible with the device and drinking (eg, sports participation) and found minimal reactions to the device [41].

Appearance

Most of the views on SCRAM's appearance were negative [25,30,37]. However, some participants described being less concerned with appearance, as they were motivated to change their alcohol consumption and were looking for help [37]. In Wang et al's study [38], the design of BACtrack was rated positively, reporting that it was lightweight and comparable to a watch.

Ease of Use

Alessi et al [23] found there were 62 adjustments required across 39 participants (N=100 participants). Of these, 56 out of 62 (90%) adjustments were because the device strap was too loose or tight; 3 out of 62 (5%) adjustments were requested by Alcohol Monitoring Systems (AMS) Inc, now better known as SCRAM Systems, and 3 out of 62 (5%) adjustments were requested by both the participants and AMS [23]. Another study reported needing to adjust, reinitialize, or replace the device 5 times in a sample of 19 participants [36]. To download SCRAM data remotely, a home phone landline or cellular signal-based modem is required. Participants who did not have one had to visit family and friends and avoid interruption of personal calls when the modem was connecting to the server to download data, thereby increasing user burden [25]. If this was also not an option for participants, data can be downloaded in-person. One study specifically asked participants to rate the ease of use of WrisTAS; the average rating was 1.2 (scale of 1-10, 1=very easy-to-use) and 26 out of 31 (84%) reported they were confident using the device after the initial session [40]. When asked about the use of Skyn compared to a breathalyzer, the results showed that Skyn was more acceptable, and the most liked features were the ease of use and design [38].

Social Perceptions

There were reports that friends and family reacted positively to SCRAM [30,36]. However, a minority reported negative judgements and noted that the ability to hide the device under clothing was appreciated. On hotter days, some chose to wear long trousers to cover the device. This was also reflected with recruitment during the hottest months of summer being difficult [25,37]. Some participants, especially those who had previous links to the criminal justice system, reported feeling embarrassed wearing the device, feared others seeing the device, and were concerned about police harassment [37]. Negative attention from others to the device in 1 study caused 2 participants to withdraw [12]. However, participants in another study rated the social discomfort of SCRAM more moderately at 4.59 (scale of 1-10, 10=extremely uncomfortable) [14]. One study, conducted in a music festival setting with healthy adults, found that most participants did not mind others seeing; some wore long socks, but most did not hide it. Generally, reactions were positive; however, 1 participant asked for the device to be removed due to concerns about not being served in bars [30].

Perceptions of Alcohol Use

A frequent comment was that wearing the device worked as a reminder and motivator to not drink alcohol [7,23,25,32,34]. Averill et al [25] using qualitative data suggested that participants believed wearing the device helped reduce drinking and this was supported by quantitative findings. Reduction in alcohol consumption was also rated as the most common personal advantage. Although 1 study found that participants reported purposefully increasing their drinking habits to provide more research data for the device, they also noted this was hard to maintain throughout [30].

Barriers, Suggestions, and Future Research

Suggested improvements by researchers and wearers included a smaller size [23,30], being waterproof, improving comfortability [12,14,23,25,38], adjustable straps [30], more notifications about data uploads from the device [38], more information about their transdermal alcohol concentration feedback [12], longer battery life [38], the use of motion or environmental sensors to corroborate output for BACtrack Skyn, and device algorithms to evaluate when deviations in recorded transdermal alcohol concentration are due to environmental factors [38].

Criminal Justice System

Goodall et al [32] conducted focus groups with serving offenders in a Scottish prison and found positive views—almost all believed there was an association between alcohol consumption and their offence. Participants reported that knowing someone was monitoring would be an incentive to reduce alcohol consumption, in addition to being constructive in knowing that there would be a consequence if they did consume alcohol. Another suggestion from the focus group was to link wearing the device to a reduced sentence.

Feasibility Measures

Device Tampering

Five studies specifically mention participants tampering with SCRAM [14,23,24,33,36]. Tamper alerts are signaled by AMS if the infrared sensor detects a deviation of 12% above or 17% below the established baseline. Of those 5 studies, 2 report very similar findings [23,24], confirmed tampering in approximately 2% of cases (105 days out of 5017 days of collected data, total participants enrolled: N=66 [24], 139 days out of 6950 days of collected data, total participants enrolled: N=100 [23]), and around half of those tamper days coincided with participants drinking. Confirmed tampers, not linked to drinking, were inadvertently caused by discomfort and subsequently repositioned. Rash et al [36] found 7 events of tampering occurred in 3 participants, most coinciding with drinking. For both studies by Alessi et al [23,24], participants were clinical alcohol treatment outpatients, and in Rash et al [36] the sample consisted of heavy drinkers attending soup kitchens.

Mathias et al [33] were concerned that some of the data suggested instances of tampering. These instances coincided with evidence of alcohol consumption (transdermal alcohol concentration>0.2) and AMS confirmed tamper events (AMS provide independent monitoring of data for the temperature sensor, an infrared sensor, and continuously conducts diagnostic tests to confirm device function). From this, they were able to retrospectively look at data and amend it for future collection.

In this study, data influenced CM; therefore, it was important to ensure withholding CM when target behavior did not occur [33]. Barnett et al [14] found that 5.5% of the data were missing due to bracelet removal due to device malfunction and tampering. AMS was able to detect specific tampering instances and the research staff were able to identify and remove specific data.

Compliance

A few studies noted 100% compliance [8,29,31,39,43]. Compliance was defined for these studies as no dropouts, no device removal, no missed appointments or assessments, and all participants wore the device for the entire intervention. Other studies mention reasons for not complying, including relocation, events incompatible with continuing (court appearance for driving under alcohol influence, incarceration, day surgery), personal discomfort, concern about employability/negative attention, and device malfunction or inconvenience [12,14,23,24,30,35,36,42]. Although all participants were asked not to remove the devices for any reason other than showering, removal while intoxicated was detected by the temperature sensor [42]. Studies found during recruitment that not everyone was willing to use the devices. Alessi et al [24] found 1 in 10 declined and Alessi et al [23] in another study found that 56 out of 595 (9.4%) participants declined due to SCRAM. Averill et al [25] found that recruitment slowed in summer, and this was suggested to be because the heat would make SCRAM more inconvenient. Compliance was high in the 4 studies with diagnosed alcohol-dependent individuals, one with 100% [8], one with 95% [24], and one with 84% [23] compliance. One does not report any participant dropouts, but device removal was detected [42].

Discussion

This review aims to assess TAS acceptability and feasibility to objectively monitor alcohol consumption in clinical and nonclinical populations. We identified 22 studies that used or investigated the attitudes and experiences of people using TAS devices. Although the available data do suggest that TAS devices are acceptable, feasible, and have the potential to monitor objective alcohol consumption data, only a few studies were conducted with clinical populations (4 out of 22 studies) or had a specific focus on acceptability and feasibility measures (8 out of 22 studies). We investigated the acceptability outcomes of wearing the TAS, including the comfort, social comfort, appearance, and ease of use of TAS. SCRAM and WrisTAS were reported to provide moderate device comfort, social comfort, and high ease of use. BACtrack was also rated highly for appearance and ease of use. However, there were also reports that SCRAM, the biggest and bulkiest device, caused skin irritation. One study [41] tried to overcome potential skin irritation by allowing participants to remove the device (WrisTAS) before physical activities. Having the freedom to adjust and remove the device for occasions such as bathing or physical activity is not typical, but as the irritation due to TASs during specific activities is one of the known issues affecting each brand of TAS at various levels, this could be a method to reduce this consequence on wearers. However, TASs require precise and secure placement for optimal data collection [44-46]; therefore, removability may only be possible after device training and participants must be trusted to replace the device. Device removability could mean that participants take off the device when consuming alcohol, place it on another sober individual, or not replace it at all. Therefore, removal of devices may not be possible in all situations.

Another uncertainty of these devices is the frequency of device malfunction and how this could impact use. If there were multiple WrisTAS malfunctions, this would not be feasible within clinical treatment or the criminal justice system. If the patient was required to return multiple times, there is an increasing burden on the device wearer and staff. Or, if there were multiple data points lost, the benefit of using this device would diminish and not be able to reliably reinforce criminal sentences or research incentives such as CM. In the included studies in this review, reported malfunctions, noise, and missing data for WrisTAS and BACtrack appeared to be much higher than those for SCRAM [9,39,40,47]. However, there are other advantages of WrisTAS and BACtrack, such as a reduced lag time and physical size and appearance [9,10,38,40]. Some individuals chose not to participate specifically due to the devices. Personal preferences for treatment options are to be expected. There were also reports of dropouts, although dropouts are typical within research. Previous research is aware of the potential discomfort [12,14,23,25,30,37]: TASs must be worn tightly on the wrist for optimal data collection and the wrist is a part of the body that is often on show—this inconvenience could be a reason for withdrawal. Although there are these concerns, many studies found high compliance, including those with alcohol-dependent individuals [7,8,23,24,29-31,35,36,39,42,43].

The only device that has the option of landline use or cellular signal–based modem is SCRAM. This may not be possible for some populations without a stable home or landline; however, for other situations where real-time data collection is not required or when regular research visits for data download is not possible, this would be very beneficial as data download can be done by the participant with no staff member present. BACtrack requires regular data downloading at least every 3 days; otherwise, data are wiped on the device. BACtrack download requires download from the device to an Apple-iPhone operating system device by Bluetooth; for some participants, this may be something they own and can be done remotely and then reviewed online by staff. However, in certain populations, only a few may already own smartphones, specifically for BACtrack, iPhones, or an iPhone operating system device with up-to-date software. One study asked participants to complete self-report web-based data collection surveys, and this was rated as easy-to-use; however, not everyone has access to the internet [40]. The newer generation TASs such as BACtrack avoid this burden by not requiring a landline. However, it needs to be considered if target users would need to be provided a smartphone as well as the TAS device.

Qualitative results suggest that wearing the device appeals as an intervention to reduce alcohol consumption due to the knowledge of someone monitoring acting as a motivator to

change their behavior [23,25,32,34]. This could be linked to and incorporate behavior change techniques within the intervention design. However, one aspect that requires further research is if alcohol-dependent patients would feel "under surveillance" by wearing this device and if this could negatively impact the trust and rapport they have with their key worker and service staff.

The potential uses of TASs include alcohol treatment and research contexts. When working with individuals currently receiving alcohol treatment or diagnosed as alcohol-dependent, the amount of alcohol consumed may be a lot greater than that investigated within laboratory studies or with healthy adults not diagnosed as alcohol-dependent. The devices may also be worn for potentially longer periods and used by individuals while heavily intoxicated. These differences could bring to light other considerations that would not be measured within laboratory studies using healthy adults and restricted alcohol consumption, under the eye of a research team. Within short-term laboratory studies with volunteers and payment, there is little reason to not comply. Device tampering is a possibility, and the likelihood of this is increased in populations who use devices within treatment and criminal justice settings. Devices typically contain temperature sensors, which can detect if the device has been removed from the skin. The ability to see this allows for discussion with patients and if appropriate, for tampered data to be removed. Testing to see if the device has been removed, placed on another individual, or any other forms of tampering are issues that could become detectable by these devices as technology advances further.

This review highlights a small number of studies investigating the acceptability and feasibility of TAS devices to objectively monitor alcohol consumption and to compare between devices by using more than one device. Given the growth in the use and appeal of this technology, further research is needed to inform interventions and policy guidance [38,48-50]. There is no standardized method for measuring the acceptability and feasibility of TASs, and there is a need for this to facilitate comparison across different devices.

There is a lack of research on the acceptability and feasibility of TASs to objectively monitor alcohol consumption in any setting, let alone within clinical or criminal justice populations. Although the available data do suggest these devices are acceptable and feasible and have the potential to capture alcohol-monitoring data, there is a need for further research within clinical populations by using robust studies outside a laboratory environment, with long-term monitoring periods. With advancements in technology and the evolution of various TASs coming to market, the focus should potentially be more on the common features of these devices rather than specific brands to better establish the potential of this type of technology. We need to further investigate how clinical populations engage with this technology and any changes in their adherence and use over extended periods of time. This can inform if and how these devices can be implemented in clinical treatment settings with or without other treatment.

Acknowledgments

This is a summary of independent research funded by the National Institute for Health and Care Research Applied Research Collaboration (NIHR ARC) South London. The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health and Social Care. CD and PD were supported by the NIHR Specialist Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King's College London. They were also supported by the NIHR Collaboration for Leadership in Applied Health Research and Care at King's College Hospital NHS Foundation Trust and the NIHR Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust. CD is supported by an NIHR Senior Investigator Award.

Authors' Contributions

EB conceptualized the work, designed the protocol, completed the systematic review and data extraction, and drafted and revised the paper. PD conceptualized the work, edited the protocol, was the third reviewer for data extraction, and revised the paper. SH was the second reviewer for data extraction and revised the paper. CD conceptualized the work, edited the protocol, and revised the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed information on the following transdermal alcohol sensor devices: SCRAM, WrisTAS, BACtrack, Milo Sensors, and Quantac.
[DOCX File , 719 KB - humanfactors_v9i4e40210_app1.docx]

Multimedia Appendix 2
Database searches, screening, and inclusion of studies for acceptability and feasibility review.
[DOCX File , 30 KB - humanfactors_v9i4e40210_app2.docx]

Multimedia Appendix 3
Database search terms.
[DOCX File , 17 KB - humanfactors_v9i4e40210_app3.docx]

Multimedia Appendix 4
Data extraction form.
[DOCX File , 14 KB - humanfactors_v9i4e40210_app4.docx]

Multimedia Appendix 5
Risk of bias Mixed Methods Appraisal Tool score for acceptability and feasibility review.
[DOCX File , 21 KB - humanfactors_v9i4e40210_app5.docx]

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Litten RZ, Bradley AM, Moss HB. Alcohol biomarkers in applied settings: recent advances and future research opportunities. Alcohol Clin Exp Res 2010 Jun;34(6):955-967. [doi: 10.1111/j.1530-0277.2010.01170.x] [Medline: 20374219]
3. Tavakoli HR, Hull M, Michael Okasinski L. Review of current clinical biomarkers for the detection of alcohol dependence. Innov Clin Neurosci 2011 Mar;8(3):26-33 [FREE Full text] [Medline: 21487543]
4. Stewart SH, Koch DG, Burgess DM, Willner IR, Reuben A. Sensitivity and specificity of urinary ethyl glucuronide and ethyl sulfate in liver disease patients. Alcohol Clin Exp Res 2013 Jan;37(1):150-155 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01855.x] [Medline: 22725265]
5. Lowe JM, McDonell MG, Leickly E, Angelo FA, Vilardaga R, McPherson S, et al. Determining ethyl glucuronide cutoffs when detecting self-reported alcohol use in addiction treatment patients. Alcohol Clin Exp Res 2015 May;39(5):905-910 [FREE Full text] [doi: 10.1111/acer.12699] [Medline: 25866234]
6. Schröck A, Thierauf-Emberger A, Schürch S, Weinmann W. Phosphatidylethanol (PEth) detected in blood for 3 to 12 days after single consumption of alcohol-a drinking study with 16 volunteers. Int J Legal Med 2017 Jan;131(1):153-160 [FREE Full text] [doi: 10.1007/s00414-016-1445-x] [Medline: 27596747]
7. Wang Y, Fridberg DJ, Leeman RF, Cook RL, Porges EC. Wrist-worn alcohol biosensors: Strengths, limitations, and future directions. Alcohol 2019 Dec;81:83-92 [FREE Full text] [doi: 10.1016/j.alcohol.2018.08.013] [Medline: 30179709]
8. Sakai JT, Mikulich-Gilbertson SK, Long RJ, Crowley TJ. Validity of transdermal alcohol monitoring: fixed and self-regulated dosing. Alcohol Clin Exp Res 2006 Jan;30(1):26-33. [doi: 10.1111/j.1530-0277.2006.00004.x] [Medline: 16433729]
9. Fairbairn CE, Kang D. Temporal Dynamics of Transdermal Alcohol Concentration Measured via New-Generation Wrist-Worn Biosensor. Alcohol Clin Exp Res 2019 Oct;43(10):2060-2069 [FREE Full text] [doi: 10.1111/acer.14172] [Medline: 31469451]
10. Marques PR, McKnight AS. Field and laboratory alcohol detection with 2 types of transdermal devices. Alcohol Clin Exp Res 2009 Apr;33(4):703-711. [doi: 10.1111/j.1530-0277.2008.00887.x] [Medline: 19170663]
11. Barrio P, Teixidor L, Andreu M, Gual A. What Do Real Alcohol Outpatients Expect about Alcohol Transdermal Sensors? J Clin Med 2019 Jun 05;8(6):795 [FREE Full text] [doi: 10.3390/jcm8060795] [Medline: 31195625]
12. Barnett NP, Tidey J, Murphy JG, Swift R, Colby SM. Contingency management for alcohol use reduction: a pilot study using a transdermal alcohol sensor. Drug Alcohol Depend 2011 Nov 01;118(2-3):391-399 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.04.023] [Medline: 21665385]
13. Dougherty DM, Karns TE, Mullen J, Liang Y, Lake SL, Roache JD, et al. Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking. Drug Alcohol Depend 2015 Mar 01;148:77-84 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.12.021] [Medline: 25582388]
14. Barnett NP, Celio MA, Tidey JW, Murphy JG, Colby SM, Swift RM. A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. Addiction 2017 Jun;112(6):1025-1035 [FREE Full text] [doi: 10.1111/add.13767] [Medline: 28107772]
15. Kilmer B, Nicosia N, Heaton P, Midgette G. Efficacy of frequent monitoring with swift, certain, and modest sanctions for violations: insights from South Dakota's 24/7 Sobriety Project. Am J Public Health 2013 Jan;103(1):e37-e43. [doi: 10.2105/AJPH.2012.300989] [Medline: 23153129]
16. Roberts E, Turley C, Piggott H, Lynch-Huggins S, Wishart R, Kerr J. Evaluation of the AAMR tagging pilot: Year 2 process evaluation findings. NatCen Social Research. 2019 Oct. URL: https://www.northyorkshire-pfcc.gov.uk/content/uploads/2020/01/Evaluation-of-the-AAMR-tagging-pilot.pdf [accessed 2022-12-14]
17. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction 2003 Dec;98 Suppl 2:1-12. [doi: 10.1046/j.1359-6357.2003.00586.x] [Medline: 14984237]
18. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res 2011 Oct;2(4):320-332 [FREE Full text] [doi: 10.1504/IJBHR.2011.043414] [Medline: 25383095]

19. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. Health Qual Life Outcomes 2010 Jul 10;8:65 [FREE Full text] [doi: 10.1186/1477-7525-8-65] [Medline: 20618978]
20. Feunekes GI, van 't Veer P, van Staveren WA, Kok FJ. Alcohol intake assessment: the sober facts. Am J Epidemiol 1999 Jul 01;150(1):105-112. [doi: 10.1093/oxfordjournals.aje.a009909] [Medline: 10400547]
21. Boniface S, Kneale J, Shelton N. Drinking pattern is more strongly associated with under-reporting of alcohol consumption than socio-demographic factors: evidence from a mixed-methods study. BMC Public Health 2014 Dec 18;14:1297 [FREE Full text] [doi: 10.1186/1471-2458-14-1297] [Medline: 25519144]
22. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
23. Alessi SM, Barnett NP, Petry NM. Experiences with SCRAMx alcohol monitoring technology in 100 alcohol treatment outpatients. Drug Alcohol Depend 2017 Sep 01;178:417-424 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.05.031] [Medline: 28709081]
24. Alessi SM, Barnett NP, Petry NM. Objective continuous monitoring of alcohol consumption for three months among alcohol use disorder treatment outpatients. Alcohol 2019 Dec;81:131-138 [FREE Full text] [doi: 10.1016/j.alcohol.2019.01.008] [Medline: 30710610]
25. Averill F, Brown TG, Robertson RD, Tchomgang A, Berbiche D, Nadeau L, et al. Transdermal alcohol monitoring combined with contingency management for driving while impaired offenders: A pilot randomized controlled study. Traffic Inj Prev 2018 Jul 04;19(5):455-461. [doi: 10.1080/15389588.2018.1448079] [Medline: 29543499]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Brobbin E, Deluca P, Hemrage S, Drummond C. Accuracy of Wearable Transdermal Alcohol Sensors: Systematic Review. J Med Internet Res 2022 Apr 14;24(4):e35178 [FREE Full text] [doi: 10.2196/35178] [Medline: 35436239]
28. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
29. Ayala J, Simons K, Kerrigan S. Quantitative determination of caffeine and alcohol in energy drinks and the potential to produce positive transdermal alcohol concentrations in human subjects. J Anal Toxicol 2009;33(1):27-33. [doi: 10.1093/jat/33.1.27] [Medline: 19161666]
30. Caluzzi G, Pennay A, Cook M, Wright C, Norman T, Kuntsche E. Transdermal monitors to assess alcohol consumption in real-time and real-life – a qualitative study on user-experience. Addiction Research & Theory 2019 Jan 09;27(4):354-361. [doi: 10.1080/16066359.2018.1530765]
31. Fairbairn CE, Bresin K, Kang D, Rosen IG, Ariss T, Luczak SE, et al. A multimodal investigation of contextual effects on alcohol's emotional rewards. J Abnorm Psychol 2018 May;127(4):359-373 [FREE Full text] [doi: 10.1037/abn0000346] [Medline: 29745701]
32. Goodall CA, Neville FG, Williams DJ, Donnelly PD. Preliminary research informing policy on remote alcohol monitoring in criminal justice: the Scottish experience. Int J Public Health 2016 Nov;61(8):865-872. [doi: 10.1007/s00038-016-0886-9] [Medline: 27624623]
33. Mathias CW, Hill-Kapturczak N, Karns-Wright TE, Mullen J, Roache JD, Fell JC, et al. Translating transdermal alcohol monitoring procedures for contingency management among adults recently arrested for DWI. Addict Behav 2018 Aug;83:56-63 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.033] [Medline: 29397211]
34. Neville FG, Williams DJ, Goodall CA, Murer JS, Donnelly PD. An experimental trial exploring the impact of continuous transdermal alcohol monitoring upon alcohol consumption in a cohort of male students. PLoS One 2013;8(6):e67386 [FREE Full text] [doi: 10.1371/journal.pone.0067386] [Medline: 23825656]
35. Norman T, Peacock A, Ferguson SG, Kuntsche E, Bruno R. Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival. Drug Alcohol Rev 2021 Nov;40(7):1112-1121. [doi: 10.1111/dar.13215] [Medline: 33174260]
36. Rash CJ, Petry NM, Alessi SM, Barnett NP. Monitoring alcohol use in heavy drinking soup kitchen attendees. Alcohol 2019 Dec;81:139-147 [FREE Full text] [doi: 10.1016/j.alcohol.2018.10.001] [Medline: 30308286]
37. Villalba K, Cook C, Dévieux JG, Ibanez GE, Oghogho E, Neira C, et al. Facilitators and barriers to a contingency management alcohol intervention involving a transdermal alcohol sensor. Heliyon 2020 Mar;6(3):e03612 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03612] [Medline: 32258468]
38. Wang Y, Fridberg DJ, Shortell DD, Leeman RF, Barnett NP, Cook RL, et al. Wrist-worn alcohol biosensors: Applications and usability in behavioral research. Alcohol 2021 May;92:25-34 [FREE Full text] [doi: 10.1016/j.alcohol.2021.01.007] [Medline: 33609635]
39. Croff JM, Hartwell ML, Chiaf AL, Crockett EK, Washburn IJ. Feasibility and reliability of continuously monitoring alcohol use among female adolescents and young adults. Drug Alcohol Rev 2021 Nov;40(7):1143-1154. [doi: 10.1111/dar.13045] [Medline: 32185847]

40. Luczak SE, Rosen IG, Wall TL. Development of a real-time repeated-measures assessment protocol to capture change over the course of a drinking episode. Alcohol Alcohol 2015 Mar;50(2):180-187 [FREE Full text] [doi: 10.1093/alcalc/agu100] [Medline: 25568142]
41. Simons JS, Wills TA, Emery NN, Marks RM. Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms. Addict Behav 2015 Nov;50:205-212 [FREE Full text] [doi: 10.1016/j.addbeh.2015.06.042] [Medline: 26160523]
42. Swift RM, Martin CS, Swette L, LaConti A, Kackley N. Studies on a wearable, electronic, transdermal alcohol sensor. Alcohol Clin Exp Res 1992 Aug;16(4):721-725. [doi: 10.1111/j.1530-0277.1992.tb00668.x] [Medline: 1530135]
43. Rosenberg M, Ludema C, Kianersi S, Luetke M, Jozkowski K, Guerra-Reyes L, et al. Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability in a pilot study. MedRxiv 2021 Feb 22:1-23. [doi: 10.1101/2021.02.17.21251959]
44. Li B, Downen RS, Dong Q, Tran N, LeSaux M, Meltzer AC, et al. A Discreet Wearable IoT Sensor for Continuous Transdermal Alcohol Monitoring - Challenges and Opportunities. IEEE Sens J 2021 Feb 15;21(4):5322-5330 [FREE Full text] [doi: 10.1109/jsen.2020.3030254] [Medline: 33746626]
45. Hill-Kapturczak N, Roache JD, Liang Y, Karns TE, Cates SE, Dougherty DM. Accounting for sex-related differences in the estimation of breath alcohol concentrations using transdermal alcohol monitoring. Psychopharmacology (Berl) 2015 Jan;232(1):115-123 [FREE Full text] [doi: 10.1007/s00213-014-3644-9] [Medline: 24923985]
46. Dougherty DM, Charles NE, Acheson A, John S, Furr RM, Hill-Kapturczak N. Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking. Exp Clin Psychopharmacol 2012 Oct;20(5):373-381 [FREE Full text] [doi: 10.1037/a0029021] [Medline: 22708608]
47. Fairbairn CE, Kang D, Bosch N. Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. Drug Alcohol Depend 2020 Nov 01;216:108205 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108205] [Medline: 32853998]
48. McKnight S, Fell CJ, Auld-Owens A. Transdermal alcohol monitoring: case studies (report DOT HS 811 603). National Academies. 2012. URL: https://trid.trb.org/view/1212685 [accessed 2022-12-14]
49. Piasecki TM. Assessment of Alcohol Use in the Natural Environment. Alcohol Clin Exp Res 2019 Apr;43(4):564-577 [FREE Full text] [doi: 10.1111/acer.13975] [Medline: 30748019]
50. Barnett NP. Alcohol sensors and their potential for improving clinical care. Addiction 2015 Jan;110(1):1-3. [doi: 10.1111/add.12764] [Medline: 25515824]

Abbreviations

AMS: Alcohol Monitoring Systems
CM: contingency management
MMAT: Mixed Methods Appraisal Tool
PRISMA-P: Preferred Reporting Items for Systematic review and Meta-Analysis for Protocols
TAS: transdermal alcohol sensor

Edited by A Kushniruk; submitted 10.06.22; peer-reviewed by M Rosenberg, M Anderson; comments to author 06.08.22; revised version received 31.08.22; accepted 07.11.22; published 23.12.22.

Please cite as:
Brobbin E, Deluca P, Hemrage S, Drummond C
Acceptability and Feasibility of Wearable Transdermal Alcohol Sensors: Systematic Review
JMIR Hum Factors 2022;9(4):e40210
URL: https://humanfactors.jmir.org/2022/4/e40210
doi:10.2196/40210
PMID:36563030

Original Paper

The Introduction of Robotics to an Outpatient Dispensing and Medication Management Process in Saudi Arabia: Retrospective Review of a Pharmacy-led Multidisciplinary Six Sigma Performance Improvement Project

Manal Al Nemari[1*], RPh; James Waterson[2*], BA, MMedEd, MHEc, RN

[1]Pharmacy Informatics and Automation, King Fahad Medical City, General Administration of Pharmaceutical Care, Ministry of Health, Riyadh, Saudi Arabia

[2]Medication Management Solutions, Medical Affairs, Becton Dickinson, Dubai, United Arab Emirates

[*]all authors contributed equally

Corresponding Author:
James Waterson, BA, MMedEd, MHEc, RN
Medication Management Solutions
Medical Affairs
Becton Dickinson
11F Blue Bay Tower
Business Bay
Dubai, 52279
United Arab Emirates
Phone: 971 0566035154
Email: james.waterson@bd.com

Abstract

Background: Outpatient pharmacy management aims for improved patient safety, improved quality of service, and cost reduction. The Six Sigma method improves quality by eliminating variability, with the goal of a nearly error-free process. Automation of pharmacy tasks potentially offers greater efficiency and safety.

Objective: The goal was to measure the impact that integration of automation made to service, safety and efficiency, staff reallocation and reorientation, and workflow in the outpatient pharmacy department. The Six Sigma problem definition to be resolved was as follows: The current system of outpatient dispensing denies quality to patients in terms of waiting time and contact time with pharmacy professionals, incorporates risks to the patient in terms of mislabeling of medications and the incomplete dispensing of prescriptions, and is potentially wasteful in terms of time and resources.

Methods: We described the process of introducing automation to a large outpatient pharmacy department in a university hospital. The Six Sigma approach was used as it focuses on continuous improvement and also produces a road map that integrates tracking and monitoring into its process. A review of activity in the outpatient department focused on non-value-added (NVA) pharmacist tasks, improving the patient experience and patient safety. Metrics to measure the impact of change were established, and a process map analysis with turnaround times (TATs) for each stage of service was created. Discrete events were selected for correction, improvement, or mitigation. From the review, the team selected key outcome metrics, including storage, picking and delivery dispensing rates, patient and prescription load per day, average packs and lines per prescription, and lines held. Our goal was total automation of stock management. We deployed 2 robotic dispensing units to feed 9 dispensing desks. The automated units were integrated with hospital information technology (HIT) that supports appointments, medication records, and prescriptions.

Results: Postautomation, the total patient time in the department, including the time interacting with the pharmacist for medication education and counseling, dropped from 17.093 to 11.812 digital minutes, with an appreciable increase in patient-pharmacist time. The percentage of incomplete prescriptions dispensed versus orders decreased from 3.0% to 1.83%. The dispensing error rate dropped from 1.00% to 0.24%. Assessed via a “basket” of medications, wastage cost was reduced by 83.9%. During implementation, it was found that NVA tasks that were replaced by automated processes were responsible for an extensive loss of pharmacist time. The productivity ratio postautomation was 1.26.

Conclusions: The Six Sigma methodology allowed for rapid transformation of the medication management process. The risk priority numbers (RPNs) for the "wrong patient-wrong medication error" reduced by a ratio of 5.25:1 and for "patient leaves unit with inadequate counseling" postautomation by 2.5:1. Automation allowed for ring-fencing of patient-pharmacist time. This time needs to be structured for optimal effectiveness.

(JMIR Hum Factors 2022;9(4):e37905) doi:10.2196/37905

KEYWORDS

inventory waste; mislabeling events; no-show returns; inventory stock levels; staff education; task realignment; outpatient; Six Sigma; medication management; medication adherence; risk; pharmacy; health care professional; dispensing; robotics; automation; pharmaceuticals; inventory

Introduction

Background

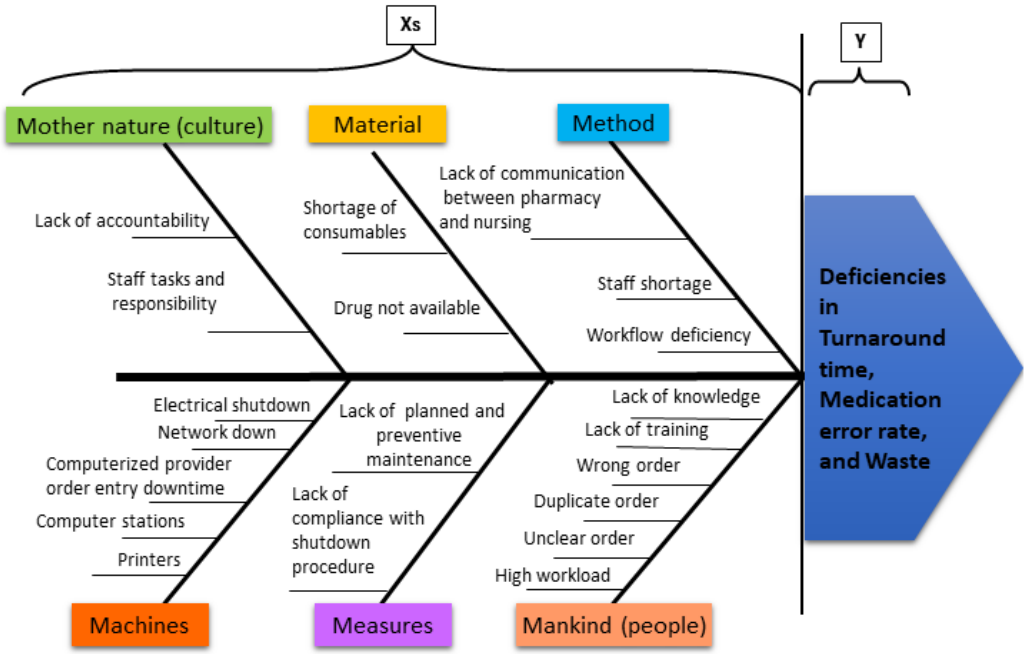
Outpatient care and outpatient pharmacy management should aim for improved patient safety, improved quality of service, and reduction in costs [1-3]. The Six Sigma method shares these same goals and aims to improve quality and reduce costs by eliminating variability, with the goal of a nearly error-free process [4]. Automation of non-value-added (NVA) pharmacy tasks that may be undertaken with greater efficiency and more safely by automated units have been identified in the literature [5], and although they will inevitably differ by facility, these are generally given as:

- Stock inputting [5]
- Inventory checks and securing of stock [5]
- Stock rotation by expiry date [6]
- Locating and picking stock [1]
- Guardianship and mandated shift-end counts for controlled substances and high-value medications [7]

King Fahad Medical City began operations in 1984, and the facility has 1200 inpatient beds and receives 500,000 outpatient visits yearly, with an increase of ~15% per year over the study period. It serves the heart of metropolitan Riyadh. Given this large and increasing workload and the increasing complexity of patient conditions being treated in the outpatient department, a systematic review of activity in the outpatient department was undertaken in 2017 as a first step in a change management process aimed at reducing to a minimum NVA pharmacist tasks, improving the patient experience, and improving patient safety. Identified issues at this point were chiefly related to the difficulties in ensuring safe dispensing due to a lack of transparency in the system and the risk of mislabeling products for dispensing, the risk of reinforcing error through labeling the product early in the medication-dispensing chain, pharmacists' value being lost in NVA tasks, the slow speed of medication picking with rising service requirements per year, a lack of integration between prescribing and dispensing requiring manual checks during picking, and substantial issues over medication availability and avoidable waste through stock expiry.

These issues were taken forward to the fuller review and building of an Ishikawa fishbone diagram of input and output deficiencies during the Six Sigma project (see Figure 1).

Figure 1. Ishikawa fishbone diagram: input and output deficiencies.



Our general philosophy during the review and going forward is that the patient experience and patient safety extend beyond metrics, such as waiting time and correct dispensing, and into the growing expectation for pharmacists to deliver effective patient education and counseling [8] and to monitor prescriptions to avoid the risk of patients receiving overwhelming or inappropriate polypharmacy [9,10].

Six Sigma is a methodology that focuses on eliminating the defects in a service or process. It is statistics based and data driven and focuses on continuous improvement [11]. In this respect, it is an excellent choice for services such as ours that have no fixed endpoint but that are rather cycles of service, quality monitoring, and improvement. Another advantage of using Six Sigma in a large and complex organization such as ours is that it produces a road map that improves tracking and monitoring as an integral part of its process.

The define-measure-analyze-improve-control (DMAIC) process is the central "spine" of the Six Sigma process. For health care, the fundamental processes first introduced into the manufacturing industry in the 1980s have been adapted to emphasize the preventive component of error reduction as this makes the process fit with other fundamental risk management processes, such as failure mode effect analysis and root-cause analysis of near misses and actual incidents [4]. The fundamentals of DMAIC for health care projects are as follows:

- Define: Delineate in detail internal customers and external customers (patients and family) and what each customer type wants and needs. Define the processes in use and their capabilities and clearly describe objectives for projected improvement efforts [12].
- Measure: Delineate the quality characteristics that would reflect improvement in customer satisfaction and process performance. Measurement also involves the creation of metrics on which the improvement efforts will be targeted.
- Analyze: Analyze data using analytical tools, such as Pareto analysis, process flow diagrams, fishbone diagrams, and statistical process control charts, to identify necessary design and process modifications for achieving customer satisfaction and performance objectives.
- Improve: Allocate resources and ring-fence them so that design and process modifications needed for improvement can be implemented rapidly and comprehensively [13].
- Control: Monitor continuously using quality management tools to ensure that the performance improvements are maintained.

Objective

The overall goal of applying a Six Sigma process improvement project plan and the objective of this study is to improve and to measure any improvement that automation, staff reallocation and reorientation, workflow change, and integration would make to service, safety, and efficiency in an outpatient pharmacy department.

The DMAIC Six Sigma change management process works most effectively when a problem statement is created to focus on an area of concern, a condition to be improved upon, or a difficulty to be eliminated [14]. We undertook a series of analyses to define the characteristics of our problem and its individual issues and to generate our problem definition:

- Supplies, inputs, process, outputs, customers (SIPOC) end-to-end actions overview (see Figure 2). The value of SIPOC analysis is that it identifies chokepoints at each stage of the input and output processes. The breakdown should have the least number of steps feasible to adequately describe the total process and ends with the internal and external stakeholders to ensure that all change is directed at improving results directly related to these individuals. In this study's SIPOC, the bulk of change was in the third stage of the process, and this directly impacted the final output in terms of medication turnaround time (TAT) and patient waiting time, ring-fenced time for counseling, and medication safety, as indicated by the fourth output stage for reported incident rates.
- Critical steps within SIPOC (see Table 1). As discussed earlier, the bulk of the activity was at the third stage of SIPOC. A more detailed process mapping of the critical steps with the average TAT and types of tasks (value-added [VA] or NVA) was created to find potential delay points and opportunities for change.
- Customer and stakeholder segmentation and identification (see Textbox 1).

Figure 2. SIPOC end-to-end mapping of medication to patient actions. CPOE: computerized provider order entry; SIPOC: supplies, inputs, process, outputs, customers.

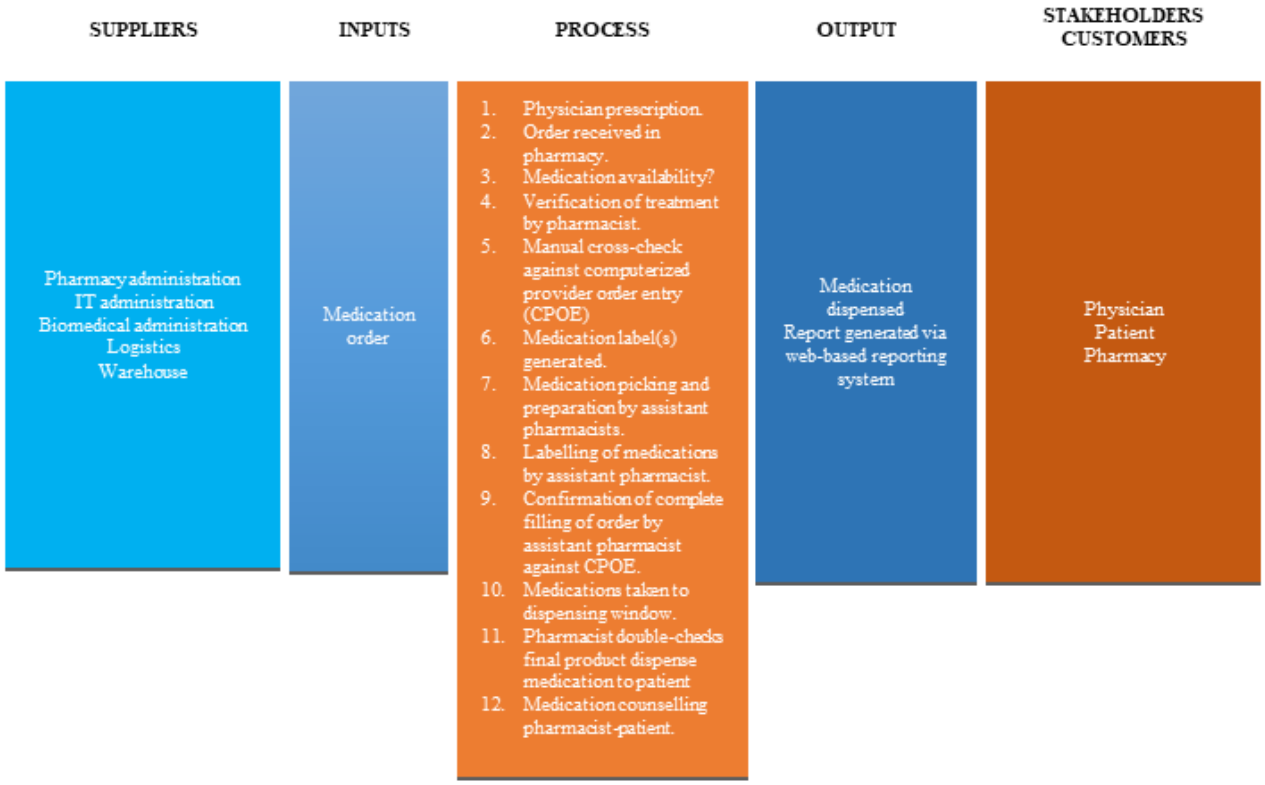


Table 1. Outpatient medication-dispensing process map before Six Sigma process application, with critical steps' mapping and TATs[a].

Step	Procedure	Projected TAT (digital minutes)	Stakeholder	VA[b]/NVA[c]	Internal/external failure	Control/inspection
1	Order a medication.	0.1	Physician	N/A[d]	• Unclear order • Physician unavailable	Outpatient clinic
2	Receive the order.	0.2	Pharmacist	VA	• Unavailability of staff • Unclear order	Support from other staff or supervisor; call doctor
3	Verify the order.	1.0	Pharmacist	VA	• Unavailability of medication	Substitute/borrow
4	Enter the medication into the CPOE[e].	1.0	Pharmacist	VA	• Unavailability of reference to check	Hardcopy reference available
5	Process the medication though the CPOE.	1.0	Pharmacist	VA	• Network downtime	Pharmacy manual backup system
6	Generate a medication label.	0.3	Pharmacist	NVA	• Network downtime	Pharmacy manual backup system
7	Pick the medication.	8.0	Pharmacist	NVA	• Unavailability of staff	Support from other staff or supervisor
8	Label the medication.	0.5	Pharmacist	NVA	• Lost label	Reprint label
9	Check the completeness of the order.	1.0	Pharmacist	NVA	• Unavailability of staff	Use technicians
10	Transport the medication to the dispensing window.	1.0	Pharmacist	NVA	• Unavailability of staff	Use technicians
11	Double-check the dispensed medication against the prescription.	1.0	Pharmacist	VA	• Unavailability of reference to check	Hardcopy reference available
12	Dispense the medication.	2.0	Pharmacist	VA	• Unavailability of staff • Refused by patient	Support from other staff or supervisor; call doctor
13	Counsel the patient and check back patient understanding.	1.0	Pharmacist	VA	• Unavailability of staff • Refused by patient	Support from other staff or supervisor; call doctor
14	The medication is received and instructions understood.	1.0	Patient	N/A	• Unavailability of staff • Refused by patient	Support from other staff or supervisor; call doctor

[a]TAT: turnaround time.

[b]VA: value-added (total projected TAT=7.2/19.1, 38%, digital minutes; number of VA tasks=7/14, 50%).

[c]NVA: non-value-added (total projected TAT=10.8/19.1, 57%, digital minutes; number of NVA tasks=5/14, 36%).

[d]N/A: not applicable (total projected TAT=1.1/19.1, 5%, digital minutes; number of N/A tasks=2/14, 14%).

[e]CPOE: computerized provider order entry.

Textbox 1. Identified customer and stakeholder segments by priority.

Internal

- Pharmacists
- Technicians
- Hospital information technology (HIT) team
- Physicians
- Warehouse stock team

External

- Patients
- Pharmaceutical vendors
- Technology vendors

The end-to-end survey of the process showed weaknesses at multiple points in terms of safety, quality, and efficiency. The survey identified the following areas of concern: medication-picking times, labeling accuracy and times, availability of medications, wastage through expiry and patient no-show, TAT for reintroduction of unused medications, and time taken in completing inventory. Of particular concern was the difficulty in constructing a viable risk management mitigation plan based on failure mode effect analysis (FMEA) that was undertaken concurrently due to a lack of transparency in the workflow and a lack of data generated by the system.

In classic FMEA planning [15] for any high-risk activity, and particularly activities with a high risk of low chance or no chance of detection of errors, the activity is broken down into individual steps, each of which can mitigate, correct, or annul any error in the previous steps. Our SIPOC did indicate a logical FMEA process to a certain degree, as steps involving the risk of error in terms of cross-checks versus the computerized provider order entry (CPOE) and picking preceded the final verification checks before the ultimate interaction with the patient by the pharmacist. However, previous steps, in particular the labeling and transport of the medication to the dispensing window by a pharmacist, were identified as adding to the risk of reinforcement of error during the final pharmacist check before release of the medication to the patient. For example, the risk during medication picking of a look-alike sound-alike (LASA) error. LASA error rates have been reported as being as high as 25.9% of all reported medication errors [16]. Furthermore, the manual application of a patient label at this stage risk influencing the pharmacist during the final check and may in fact risk reinforcing the chance of error rather than averting it [17,18].

It was also noted during our review that the time spent with the patient for medication counseling was potentially too short to allow for optimal counseling and education [19]. The patient-pharmacist time was also likely to be “squeezed” by the need to manage patient volume and to handle increasing transaction rates per day [20].

We wanted our medication chain to have improved detection, and decreased risk, of error at each stage of the medication chain. FMEA risk priority numbers (RPNs) were calculated for all perceived risks in day-to-day operations of the outpatient pharmacy department. In this study, we focused on the 2 conditions laid out earlier, as they are central to patient safety but also reflect possible efficiency gains that could be obtained by introducing automation. The FMEA RPNs (severity score [SS] × probability score [PS] × detectability score [DS]) were, respectively:

Misinterpretation of prescription-wrong patient-wrong medication error:

$$SS \times PS \times DS = 6 \times 3 \times 7 = 126$$

Patient leaves unit with inadequate counseling:

$$SS \times PS \times DS = 4 \times 5 \times 3 = 60$$

The high PS for inadequate counseling was based on the degree of variability in patient service time, discussed later. Our scores overall were in line with risk calculations made in other centers reviewing manual dispensing systems for outpatient departments [3]. See Multimedia Appendix 1 for a fuller description of the process and improvement criteria according to the Institute for Healthcare Improvement [21].

These analyses allowed for a Six Sigma problem statement to be generated:

> *The current system of outpatient dispensing denies quality to patients in terms of waiting time and contact time with pharmacy professionals, incorporates risks to the patient in terms of mislabeling of medications and incomplete dispensing of prescriptions, and is potentially wasteful in terms of time and resources.*

Methods

Study Design

The study lasted 20 months (April 2019-December 2020), with a go-live for the automated pharmacy after 3 months (July 2019).

To assess the impact of changes brought about by automation, task reassignment, and workflow restructuring, it was necessary to establish further metrics; we reviewed our Six Sigma change process and reviewed the steps delineated in Table 1 and its process map analysis of discrete events with an average TAT for each step, with a risk of delay by cause for each stage (Table

2). Going forward, the TAT was calculated from multiple observations at each step of the process.

It was possible to directly review causes for protracted TATs and frequent delays. Given the substantial number of stakeholders and multiple data points, an Ishikawa fishbone cause-and-effect diagram was built to assist the team of internal stakeholders in brainstorming to capture the sources of process variation, to drill down for the causes of delay, to investigate each detrimental effect, and to determine correctable and improvable causes [22].

The discrete events identified were selected for correction, improvement, or mitigation. The final task during our appraisal was to return to the TAT and delay assessment and, in conjunction with the Ishikawa exercise, to decide on the final variables and derived metrics that would indicate, most accurately, reliably, and appropriately, the effect of any changes we put in place. The team, supported by the facility's analytics department, selected 11 key measurable outcome variables and derived metrics (see Table 3). The review also aided with team selection as we identified process variations and chokepoints hindering improvement (see Figure 1) and could recruit personnel directly involved at these points into the project team.

Table 2. Process map analysis for average TAT[a] and delay postautomation.

Step	Procedure	Projected TAT (digital minutes)	Stakeholder	VA[b]/NVA[c]	Internal/external failure	Control/inspection
1	Order a medication.	0.1	Physician	N/A[d]	• Unclear order • Physician unavailable	Outpatient clinic
2	Receive the order.	0.2	Pharmacist	VA	• Unavailability of staff • Unclear order	Support from other staff or supervisor; call doctor
3	Verify the order.	1.0	Pharmacist	VA	• Unavailability of medication	Substitute/borrow
4	Enter the medication into the CPOE[e].	0.1	Integration	N/A	• Unavailability of reference to check	Hardcopy reference available
5	Process the medication though the CPOE.	0.1	Integration	N/A	• Network downtime	Pharmacy manual backup system
6	Generate a medication label.	0.1	Automation/integration	N/A	• Network downtime	Pharmacy manual backup system
7	Pick the medication.	1.0	Automation/integration	N/A	• Unavailability of staff	Support from other staff or supervisor
8	Label the medication.	0.1	Automation/integration	N/A	• Lost label	Reprint label
9	Check the completeness of the order.	0.1	Automation/integration	N/A	• Unavailability of staff	Use technicians
10	Transport the medication to the dispensing window.	0.2	Automation/integration	N/A	• Unavailability of staff	Use technicians
11	Double-check the dispensed medication against the prescription.	1.0	Pharmacist	VA	• Unavailability of reference to check	Hardcopy reference available
12	Dispense the medication.	1.8	Pharmacist	VA	• Unavailability of staff • Refused by patient	Support from other staff or supervisor; call doctor
13	Counsel the patient and check back patient understanding.	5.0	Pharmacist	VA	• Unavailability of staff • Refused by patient	Support from other staff or supervisor; call doctor
14	The medication is received and instructions understood.	1.0	Patient	N/A	• Unavailability of staff • Refused by patient	Support from other staff or supervisor; call doctor

[a]TAT: turnaround time.

[b]VA: value-added (total projected TAT=9.0/11.8, 76%, digital minutes; number of VA tasks=5/14, 56%).

[c]NVA: non-value-added.

[d]N/A: not applicable (total projected TAT=2.8/11.8, 24%, digital minutes; number of N/A tasks=9/14, 44%).

[e]CPOE: computerized provider order entry.

Table 3. Key measurable outcome indicators.

Variable and derived metric	Type of metric	Assessment methods	Pre-reengineering metrics	Postreengineering metrics	Inferred metric
Patient waiting time from presentation to departure with correct medications and appropriate counseling	Outcome indicator	• Total time in department – time interacting with pharmacist for medication education and counseling	• 17.093	• 11.812	Improvement in adherence, reduction in adverse drug events (ADEs) at home
Completeness of all dispensed prescriptions	Outcome indicator	• Percentage of complete prescriptions dispensed versus orders • Error rate per 1000 items dispensed	• Data of 6 months at initial and immediate stages of automation takeover (using automated pharmacy systems to account for lack of data prerobotic pharmacy) • Mean 3.0% (SD 5.07%) • Orders sampled (N=14,991)	• Data of 6 months following go-live (1) and 6 months of system stabilization (2) • 1: mean 2.99% (SD 2.82%) • 2: mean 1.83% (SD 0.99%)	Pre-post availability of medications in percentage inventory transparency
Accuracy of all dispensed medications	Outcome indicator	• Medication error reporting via incident reporting system (one year pre- and post-engineering via web-based incident reporting system) • In-out discrepancies/day • Number of mislabeling events caught • Data of 1 year pre- and postreengineering	• Mean 1.00% (SD 0.279%)	• Mean 0.24% (SD 0.299%)	Workflow FMEA[a] capacity to prevent error reaching patient
FMEA RPN[b] for "misinterpretation of prescription-wrong patient-wrong medication error" reduced from the initial score of 126	Outcome indicator	• Recalculate FMEA RPN postreengineering	• SS[c] × PS[d] × DS[e] = 6 × 3 × 7 = 126	• SS × PS × DS = 6 × 1 × 4 = 24 • PS reduced by a ratio of 4:1 (1.0%:0.24%) • DS reduced as more complete information and transparency of workflow increased	Safety of process
FMEA RPN for "patient leaves unit with inadequate counseling" reduced from the initial score of 60	Outcome indicator	• Recalculate FMEA RPN postreengineering • Pharmacist spends more time for consultation in digital minutes or % per interaction	• SS × PS × DS = 4 × 5 × 3 = 60	• SS × PS × DS = 4 × 2 × 3 = 24 • 5 digital minutes of time counseling the patient ring-fenced compared to 1 digital minute in preautomation; however, time still being structured for optimal effectiveness	Reduced likelihood of unscheduled return to hospital caused by medication nonadherence

Variable and derived metric	Type of metric	Assessment methods	Pre-reengineering metrics	Postreengineering metrics	Inferred metric
Pharmacist deployment to VA[f] vs NVA[g] tasks	Outcome indicator	• NVA tasks by employed digital minutes per item dispensed × average number of items/year dispensed • Time not including maintenance tasks and manual inventory	• Total 17.093 digital minutes per prescription • VA=6.44 digital minutes per item • NVA=9.67 digital minutes per prescription • N/A[h]=0.98 digital minutes per prescription • NVA time=mean 28,831 (SD 2992.60) new prescriptions/month × 9.67 × 12=3,345,549.2 digital minutes=2323 days 7 hours/year (24-hour service) • 1 FTE[i]=225 days/year=10.33 FTEs	• Total 11.812 digital minutes per item • NVA=0 • VA=10.247 digital minutes per item • N/A=1.565 digital minutes per item • NVA time=0	Ability to apply human resources to VA, patient-centered activities
Prescriptions filled	Process quality indicator	• Items/day/year	• Data of 1 year prereengineering≈7824 items/day≈2,856,000 items/year	• Data of 1 year postreengineering≈9884 items/day≈3,607,400 items/year	Fluctuation management, pharmacist productivity (productivity ratio=1.26)
Waste through expiry, loss, or failure to reintroduce into the system	Outcome indicator	• Calculated by cost of a “basket” of 8 diverse prescriptions • Percentage waste change postreengineering (Table 4)	• US $124,592.24 lost to waste (Table 4)	• US $20,017.98 lost to waste • 83.9% reduction in waste across basket (Table 4)	Reduction of expired stock loss: more efficient, leaner stock (lower stock level, lower cash binding)
Staff education on automated processes and workflow changes, and direction on use of freed-up time: polypharmacy, counseling techniques	Process quality indicator	• Directed in-service education time/month	• In-service teaching time devoted to workflow per month=mean 825 (SD 375) digital minutes	• In-service teaching time devoted to use of freed-up time per month=mean 637.5 (SD 341.64) digital minutes	Move pharmacists from functional roles to innovative roles

[a]FMEA: failure mode effect analysis.

[b]RPN: risk priority number.

[c]SS: severity score.

[d]PS: probability score.

[e]DS: detectability score.

[f]VA: value-added.

[g]NVA: non-value-added.

[h]N/A: not applicable.

[i]FTE: full-time employee.

Table 4. "Basket" of 8 prescription medications reviewed for costs of waste by expiry or incorrect storage over 1 year.

Medication name and dose	Unit value (US $)	Unit waste pre-reengineering	Cost waste pre-reengineering (US $)[a]	Unit waste postreengineering	Cost waste postreengineering (US $)[b]
Everolimus 10 mg	151.88	630	95,681.25	90	13,668.75
Isoniazid 10 mg/mL	65.41	9	588.67	5	327.04
Lipase 5000 international units (IU) + amylase 3600 IU + protease 200 IU granules	59.40	215	12,771.00	30	1782.00
Apomorphine 10 mg/mL	45.36	300	13,608.00	85	3855.60
Cetrorelix 0.25 mg injection	33.40	9	300.62	5	167.01
Tacrolimus 0.03% cream	14.85	55	816.75	6	89.10
Methotrexate 7.5 mg/0.15 mL syringe	12.96	34	440.64	3	38.88
Cefdinir 125 mg/5 mL suspension	8.96	43	385.31	10	89.61

[a]Total cost waste pre-reengineering=US $124,592.24.

[b]Total cost waste postreengineering=US $20,017.99.

The rationale for the selection of variables was that they directly address the key areas of the problems and questions we were trying to solve, as identified in our Six Sigma process mapping and Ishikawa deconstruction, and they also work across the spectrum of a complex of issues that are interrelated and interdependent: this was a fundamental reason for selecting Six Sigma as our change methodology, as improvements in 1 variable could impact multiple metrics and the interplay between them is well described through Six Sigma with its identification of chokepoints to improvement. The variables identified for metric development were the risk of dispensing errors, the risk of the system reinforcing the error risk rather than mitigating it, the question of how much transparency the system has, the loss of staff value, the patient experience, maintaining adequate throughput in the system to allow for ring-fenced pharmacist-patient time, ensuring medication availability through integration, and reducing medication loss through expiry/misplacing. The purpose of the derived metrics was that they triangulate with data the variables we wanted to investigate, which are difficult to address directly (Table 5).

Table 5. Selected variables and derived metrics with rationales.

Variables	Derived metrics addressing variables	Selection rationale for variables and derived metrics
Patient experience, ring-fenced pharmacist-patient time, staff value and use	Patient waiting time from presentation to departure with correct medications and appropriate counseling	Indirect measurement of patient satisfaction and experience (linked to waiting time [1]), indirect indication of probability of ring-fenced pharmacist-patient time being maintained
Risk of dispensing error, dispensing throughput, medication loss, system transparency	Completeness of all dispensed prescriptions	Direct measurement of complete orders, indirect assessment via incident reporting and root-cause analysis of medication not being available
Risk of dispensing error, system transparency	Accuracy of all dispensed medications	Direct measurement of error rate via an incident reporting system, FMEA[a] indirect assessment of error detectability via FMEA scoring pre- and postautomation
Risk of dispensing error, system transparency	FMEA RPN[b] for "misinterpretation of prescription-wrong patient-wrong medication error" reduced from the initial score of 126	Change in RPN calculated from PS[c] and DS[d]
Ring-fenced pharmacist-patient time, system transparency, staff value and use	FMEA RPN for "patient leaves unit with inadequate counseling" reduced from the initial score of 60	Change in RPN calculated from PS and DS
Ring-fenced pharmacist-patient time, system transparency, staff value and use	Pharmacist deployment to VA[e] vs NVA[f] tasks	Semidirect calculation derived from calculation of time for NVA tasks eliminated by automation against total unit throughput in medications dispensed
Dispensing throughput, system transparency, staff value and use	Prescriptions filled	Direct measurement of ability to meet demand and ability to measure volume handled, indirect measurement of how successfully automation and human systems are interacting
Medication loss	Waste through expiry, loss, or failure to reintroduce into the system	Direct and frequent inventory of medication stock and waste, expired medications counted, derived calculation of misplaced medications
Ring-fenced pharmacist-patient time, patient experience, staff value and use	Staff education on automated processes and workflow changes, and direction on use of freed-up time: polypharmacy, counseling techniques	Indirect indication of staff engaging in counseling and in patient engagement

[a]FMEA: failure mode effect analysis.

[b]RPN: risk priority number.

[c]PS: probability score.

[d]DS: detectability score.

[e]VA: value-added.

[f]NVA: non-value-added.

Our review of the materials and solutions needed for the automated "heart" of our reengineering of the process and management of the outpatient pharmacy was guided by a review of the literature [3,5]. Technology selection in terms of required storage, picking, and delivery rates was made through a review of 2017-2019 pack-dispensing rates (mean 292,662, SD 34,301 packs/month; mean 10,452, SD 1225 packs/day) for a mean patient load of 42,663 (SD 2992) patients/month, or 1524 (SD 107) patients/day.

As noted before, we experienced an increase of ~15% patients/year using the service, and we planned for this required extra capacity.

The outpatient pharmacy operated, both pre- and postautomation, a 24-hour service, with peak times from 9:00 a.m. to noon and from 4:00 p.m. to 6:00 p.m. This was important to note for planning as no restocking could be expected to take place during these periods, with both robotic units dedicated to meet the dispensing demand.

Our goal was to achieve as extensive an automation of the processes of stock management as possible; therefore, we investigated systems with semi- and fully automatic input, and this was planned to take place during low-patient-volume hours at a minimum rate of 350-500 packs/hour input.

BD Rowa Vmax 160 hardware was selected based on the aforementioned criteria for picking and input speed and positive integration attributes. Two machines were purchased, each with dimensions of 7.4 m length × 1.6 m width × 2.9 m height. Each unit has a capacity for ~12,500 medications, with a potential high-density storage capacity of 18,300-20,100 medications. The external architecture serves 9 dispensing desks via spiral chutes, fed by 2 unidirectional belts with feed gates, serviced by 1 bidirectional belt feeding from 4 exit points of the 2 robot

picking units. Given the substantial volumes of medications dispensed per day in our outpatient department, and to achieve the higher rate of medication picking and dispensing that we required to meet our targets, we opted not to connect the 2 units but instead to stock both units fully with required medications. Connecting the units has some advantages in that if 1 robotic unit is operating at a faster rate for picking, it can take over a greater load and dispense more of the served dispensing desks. However, this crossover of activity between units can interrupt picking and create small, but important, breaks in dispensing activity.

Refill of the robotic units was via an electronic refill system that triggers stock requests versus par levels for each medication held against a continuous inventory and consumption check.

The project took an open approach to integration with existing hospital information technology (HIT), as planning was in place for replacement of the information environment of 2017 with an organization-wide conversion to an Epic Systems Corporation electronic medical record (EMR) with integration into the EMR of all services and health information and EMR support for appointments, medication records, and prescription by 2022.

FutureGate Pharmaflow architecture was chosen for initial and ongoing integration of the robotic pharmacy. This solution can interface via Health Level Seven (HL7) and is therefore flexible enough to operate through a changing HIT environment.

Where extra notation and extra labeling at medication input were required for picking by the robotic pharmacy, these barcodes were generated from the Pharmaflow formulary manager, although the gradual introduction of the Global Trade Item Number (GTIN) for medications over the go-live and study period reduced this need somewhat. Whole-box dispensing was used.

The Pharmaflow solution was also chosen based on ensuring continuity of service. The solution's VMware that supports the robotic pharmacy's interoperability is entirely within the facility's firewall, and there are no requirements for inbound ports to be opened. Urgent maintenance requests are managed by a remote access broker via a Secure Sockets Layer (SSL) over Transmission Control Protocol (TCP) port 443 (outbound rule only). Virtual proxy network (VPN) access is initiated by our facility if access by vendor engineers is required for remote server maintenance. The user identification and the date and time of any action are electronically logged.

The question of wastage of medications and whether this reduces following the implementation of automation was addressed by a retrospective review of manual inventories for loss through expiry or incorrect storage of a basket of 8 diverse medications pre-reengineering and a review of automated reports postreengineering for these same medications (see Table 4). The medications were selected to be representative of a wide selection of the types of medication we carry in our outpatient inventory, from slow-moving items with small-inventory-volume items to commonly used medications with larger holding stock volumes, specialist items, and parenteral and enteral products. Higher-cost items were not used, as their single-item value could skew the inventory costs savings and they were subject to individual stock control measures.

This targeted approach was taken as data in the prephase, due to the difficulties in undertaking a systemwide inventory, were limited.

Ethical Considerations

As data are continually collected from medical devices across our facility, and all nursing and medical staff are aware of ongoing collection and analysis of actual and near-miss events, no formal consent was required from the ethics committee of King Fahad Medical City for data gathering.

Study Procedure

The data recorded for analysis were patient-anonymized for hospital number, gender, name, date of birth, or other identifiable material. All employees active in the outpatient unit were informed of the purpose of the study and the data collection taking place. Furthermore, the change management process and attendant data gathering are a consistent part of the facility's process of Joint Commission International (JCI) quality improvement and zero-harm targets. According to facility protocols, the pharmacy department owns all medication management automation data and is recognized as the lead department for medication safety. BD Clinical and FutureGate Global Customer Services were engaged to optimize the automated solution. BD Medical Affairs was requested to undertake a deeper analysis of the data. The medical affairs department of BD operates as a distinct arm outside of the commercial operations of the company.

Inclusion Criteria

All formulary tablet and capsule orders were dispensed via the outpatient pharmacy as original-pack medications either as newly prescribed or as represcribed items; this included whole-box brand name medications entered into the system via manufacturers' barcode identification and whole-box generic medications entered into the system via facility-applied barcode identifiers. Ready-for-use suspensions and suspensions requiring reconstitution were included, as were enzyme supplement packs in granular form. Prefilled medication syringes and powder and solvent injection kits were included as packs and individual items according to prescriptions.

Exclusion Criteria

The study did not include unit-dose medications for suspensions, tablets, or capsules. Patient-named medication and blister packs were excluded.

Results

Summary weekly reports from preautomation showed mean and median service times slightly below the projected TAT of 19.1 digital minutes for 75% of patients; however, this was ~47 digital minutes for 95% of patients. Postautomation results from early 2020 showed that the mean average TAT fell to 11.8 digital minutes, with 88% of patients served within 20 digital minutes (see Table 6).

Table 6. Sample prescription dispensing TATs[a] pre- and 3 months postautomation.

Sample	Preautomation	3 months postautomation
Mean (SD)	17.093 (5.743)	11.812 (3.821)
Variance	32.977	14.597
Skewness	1.26864	0.840425
Kurtosis	2.21838	0.271411
N	53	53
Minimum	8.128	6.456
1st quartile	13.619	9.026
Median	15.829	11.096
3rd quartile	20.097	14.029
Maximum	36.635	21.783
95% CI for mean (SD)	15.510-18.676 (4.280-7.105)	10.759-12.865 (3.207-4.727)
95% CI for median	14.702-16.987	9.599-12.637

[a]TAT: turnaround time.

Discussion

Principal Findings

The Six Sigma methodology allowed for rapid transformation of the medication management process. The RPN for "wrong patient-wrong medication error" reduced by a ratio of 5.25:1 and for "patient leaves unit with inadequate counseling" postautomation by 2.5:1. As per the Institute for Healthcare Improvement guidelines [21], a 50% (2:1) reduction in the RPN indicates a successful FMEA process. The percentage of incomplete prescriptions dispensed versus orders decreasing from 3.0% to 1.83% is in line with previous studies of risk analysis [3]. The dispensing error rate drop from 1.00% to 0.24% is perhaps a reflection of the change in workflow with a reduction in the risk of dispensing from "upstream" labeling errors in the pre-reengineering process.

The FMEA detectability scores are a continuing concern for counseling, although automation has allowed for ring-fencing of patient-pharmacist time. The difficulty is, of course, quantifying the effectiveness of counseling. A structured approach to patient education and the use of the "teach back" methodology may allow us to quantify the effectiveness of patient counseling more completely. It is notable that the 95% CI for the median time for medication dispensing carries a narrower dispersal of times (9.599-12.637 vs 14.702-16.987), and this reduction in variability, as well as the reduced time overall, gives us more confidence for planning for consistent structured pharmacist-patient counseling time. The extra time has been created from a reduction in NVA pharmacist tasks and reduced patient waiting time.

The evolving EMR project may be of assistance in assessing the impact of this increased patient-pharmacist time, as it can help us determine the volume of admissions related to nonadherence to medication or incorrect medication usage at home by patients. Small-scale studies have indicated that in Saudi Arabia, hospital admissions due to medication errors at home account for 14.7% of emergency room admissions, with failure to take or receive medication being responsible for 47.2% of these presentations [23].

Personnel were a central component of the Ishikawa diagram of deficiencies, and it is notable how much NVA task time was uncovered by the change to automation in terms of full-time employee (FTE) time devoted to these types of tasks. The substantial FTE time saved will not lead to a reduction in workforce; redeployment and reorientation of staff through education for more patient-directed activity are ongoing, and given that we expect ongoing increases of ~15% patients/year using the service, we recognize the need for extra capacity in all areas of the department.

We undertook a series of analyses to define the characteristics of our problem, and 1 major problem we encountered was that many of the manual activities we undertook preautomation generated no data or data that were hard to obtain and appraise. With automation, a great deal of "passive" data collection takes place, giving improved transparency to the system.

For example, the selection of a "basket" of medications for wastage review was required due to a lack of data preautomation, and the cost saving calculated from this was substantial. Postautomation, with a dynamic inventory, we can extend these reviews across the entire stock held within the robotic pharmacy. Other studies of robotic pharmacy installations have shown a return on investment (ROI) within 3.5-3.75 years [1,6], with reduced wastage as a significant component of this return.

Conclusion

Our data indicate that the efficiency and safety of the system are improving with time. We believe that these ongoing improvements are related to staff having "learned" the technology and becoming increasingly proactive in its use and being able to use the new systems more effectively, as well as exploiting opportunities presented by automation. Our initial Six Sigma problem statement included the issue that our system

was “potentially wasteful in terms of time and resources.” A key aspect of the reengineering we undertook is that we can more clearly identify where these potential losses are and more accurately target them.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of Institute for Healthcare Improvement implementation of failure mode effect analysis processes, with selected component processes.
[DOCX File , 14 KB - humanfactors_v9i4e37905_app1.docx]

References

1. Momattin H, Arafa S, Momattin S, Rahal R, Waterson J. Robotic pharmacy implementation and outcomes in Saudi Arabia: a 21-month usability study. JMIR Hum Factors 2021 Sep 01;8(3):e28381 [FREE Full text] [doi: 10.2196/28381] [Medline: 34304149]
2. Batson S, Herranz A, Rohrbach N, Canobbio M, Mitchell SA, Bonnabry P. Automation of in-hospital pharmacy dispensing: a systematic review. Eur J Hosp Pharm 2021 Mar;28(2):58-64 [FREE Full text] [doi: 10.1136/ejhpharm-2019-002081] [Medline: 32434785]
3. Rodriguez-Gonzalez C, Herranz-Alonso A, Escudero-Vilaplana V, Ais-Larisgoitia M, Ribed-Sanchez A, Tovar-Pozo M, et al. DD-012 A risk analysis method to evaluate the impact of robotic dispensing on patient safety. Eur J Hosp Pharm 2016 Feb 14;23(Suppl 1):A109.2-A1A110. [doi: 10.1136/ejhpharm-2016-000875.247]
4. Niñerola A, Sánchez-Rebull MV, Hernández-Lara AB. Quality improvement in healthcare: Six Sigma systematic review. Health Policy 2020 Apr;124(4):438-445. [doi: 10.1016/j.healthpol.2020.01.002] [Medline: 32156468]
5. Berdot S, Korb-Savoldelli V, Jaccoulet E, Zaugg V, Prognon P, Lê LMM, et al. A centralized automated-dispensing system in a French teaching hospital: return on investment and quality improvement. Int J Qual Health Care 2019 Apr 01;31(3):219-224 [FREE Full text] [doi: 10.1093/intqhc/mzy152] [Medline: 30007301]
6. Mathy C, Pascal C, Fizesan M, Boin C, Délèze N, Aujoulat O. Automated hospital pharmacy supply chain and the evaluation of organisational impacts and costs. Supply Chain Forum: Int J 2020 Jul 02;21(3):206-218. [doi: 10.1080/16258312.2020.1784687]
7. Portelli G, Canobbio M, Bitonti R, Della Costanza C, Langella R, Ladisa V. The impact of an automated dispensing system for supplying narcotics in a surgical unit: the experience of the National Cancer Institute Foundation of Milan. Hosp Pharm 2018 Sep 05;54(5):335-342. [doi: 10.1177/0018578718797265]
8. Albanese NP, Rouse MJ, Schlaifer M. Scope of contemporary pharmacy practice: roles, responsibilities, and functions of pharmacists and pharmacy technicians. J Am Pharm Assoc 2010 Mar;50(2):e35-e69. [doi: 10.1331/japha.2010.10510]
9. Hilmer SN, Gnjidic D. The effects of polypharmacy in older adults. Clin Pharmacol Ther 2009 Jan;85(1):86-88. [doi: 10.1038/clpt.2008.224] [Medline: 19037203]
10. Garcia-Caballos M, Ramos-Diaz F, Jimenez-Moleon JJ, Bueno-Cavanillas A. Drug-related problems in older people after hospital discharge and interventions to reduce them. Age Ageing 2010 Jul;39(4):430-428. [doi: 10.1093/ageing/afq045] [Medline: 20497947]
11. de Koning H, Verver JPS, van den Heuvel J, Bisgaard S, Does RJMM. Lean Six Sigma in healthcare. J Healthc Qual 2006;28(2):4-11. [doi: 10.1111/j.1945-1474.2006.tb00596.x] [Medline: 16749293]
12. DelliFraine JL, Wang Z, McCaughey D, Langabeer JR, Erwin CO. The use of Six Sigma in health care management: are we using it to its full potential? Qual Manag Health Care 2014;23(4):240-253. [doi: 10.1097/QMH.0000000000000039] [Medline: 25260101]
13. Van den Heuvel J, Does RJJM. Dutch hospital implements Six Sigma. Six Sigma Forum Magazine 2004;4(2):11-14.
14. Bryman A. The research question in social research: what is its role? Int J Soc Res Methodol 2007 Feb;10(1):5-20. [doi: 10.1080/13645570600655282]
15. Stamatis D. Failure Mode and Effects Analysis: FMEA from Theory to Execution. Milwaukee, WI: ASQC Quality Press; 1995.
16. National Pharmacy Association. NPA Medication Safety Update (MSO Report) Quarter 4 2019 (England). 2019. URL: https://www.npa.co.uk/news-and-events/news-item/npa-medication-safety-update-mso-report-quarter-4-2019-england/ [accessed 2022-09-22]
17. Waterson J, Al-Jaber R, Kassab T, Al-Jazairi AS. Twelve-month review of infusion pump near-miss medication and dose selection errors and user-initiated "good save" corrections: retrospective study. JMIR Hum Factors 2020 Aug 11;7(3):e20364 [FREE Full text] [doi: 10.2196/20364] [Medline: 32667895]

18. Institute for Safe Medication Practices. Independent Double Checks: Worth the Effort if Used Judiciously and Properly. 2019. URL: https://www.ismp.org/resources/independent-double-checks-worth-effort-if-used-judiciously-and-properly [accessed 2022-09-22]
19. Oñatibia-Astibia A, Malet-Larrea A, Larrañaga B, Gastelurrutia M, Calvo B, Ramírez D, et al. Tailored interventions by community pharmacists and general practitioners improve adherence to statins in a Spanish randomized controlled trial. Health Serv Res 2019 Jun 07;54(3):658-668 [FREE Full text] [doi: 10.1111/1475-6773.13152] [Medline: 30957240]
20. Yang S, Kim D, Choi HJ, Chang MJ. A comparison of patients' and pharmacists' satisfaction with medication counseling provided by community pharmacies: a cross-sectional survey. BMC Health Serv Res 2016 Apr 14;16(1):131 [FREE Full text] [doi: 10.1186/s12913-016-1374-x] [Medline: 27080704]
21. Institute for Healthcare Improvement. Failure Modes and Effects Analysis (FMEA) Tool. URL: https://tinyurl.com/26pkrbpv [accessed 2022-09-22]
22. American Society for Quality. Cause Analysis Tools: Fishbone Diagram. 2009. URL: http://www.asq.org/learn-about-quality/cause-analysis-tools/overview/fishbone.html [accessed 2022-08-20]
23. Al-Olah Y, Al Thiab K. Admissions through the emergency department due to drug-related problems. Ann Saudi Med 2008;28(6):426-429. [doi: 10.4103/0256-4947.51671]

Abbreviations

CPOE: computerized provider order entry
DMAIC: define-measure-analyze-improve-control
DS: detectability score
EMR: electronic medical record
FMEA: failure mode effect analysis
HIT: hospital information technology
LASA: look-alike sound-alike
NVA: non-value-added
PS: probability score
RPN: risk priority number
SIPOC: supplies, inputs, process, outputs, customers
SS: severity score
TAT: turnaround time
VA: value-added

Edited by A Kushniruk; submitted 11.03.22; peer-reviewed by YM Huang, S Abideen, D Pessoa; comments to author 01.04.22; revised version received 11.08.22; accepted 11.09.22; published 11.10.22.

Please cite as:
Al Nemari M, Waterson J
The Introduction of Robotics to an Outpatient Dispensing and Medication Management Process in Saudi Arabia: Retrospective Review of a Pharmacy-led Multidisciplinary Six Sigma Performance Improvement Project
JMIR Hum Factors 2022;9(4):e37905
URL: https://humanfactors.jmir.org/2022/4/e37905
doi:10.2196/37905
PMID:36222805

Original Paper

Clinicians' Perceptions of an Artificial Intelligence–Based Blood Utilization Calculator: Qualitative Exploratory Study

Avishek Choudhury[1], PhD; Onur Asan[2], PhD; Joshua E Medow[3], MD, PhD

[1]Industrial and Management Systems Engineering, Benjamin M Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States

[2]Systems Engineering, School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

[3]Neurocritical Care, Neurosurgery, Pathology, and Biomedical Engineering, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

Corresponding Author:
Avishek Choudhury, PhD
Industrial and Management Systems Engineering
Benjamin M Statler College of Engineering and Mineral Resources
West Virginia University
1306 Evansdale Drive
PO Box 6107
Morgantown, WV, 26506-6107
United States
Phone: 1 3042939431
Email: avishek.choudhury@mail.wvu.edu

Abstract

Background: According to the US Food and Drug Administration Center for Biologics Evaluation and Research, health care systems have been experiencing blood transfusion overuse. To minimize the overuse of blood product transfusions, a proprietary artificial intelligence (AI)–based blood utilization calculator (BUC) was developed and integrated into a US hospital's electronic health record. Despite the promising performance of the BUC, this technology remains underused in the clinical setting.

Objective: This study aims to explore how clinicians perceived this AI-based decision support system and, consequently, understand the factors hindering BUC use.

Methods: We interviewed 10 clinicians (BUC users) until the data saturation point was reached. The interviews were conducted over a web-based platform and were recorded. The audiovisual recordings were then anonymously transcribed verbatim. We used an inductive-deductive thematic analysis to analyze the transcripts, which involved applying predetermined themes to the data (deductive) and consecutively identifying new themes as they emerged in the data (inductive).

Results: We identified the following two themes: (1) workload and usability and (2) clinical decision-making. Clinicians acknowledged the ease of use and usefulness of the BUC for the general inpatient population. The clinicians also found the BUC to be useful in making decisions related to blood transfusion. However, some clinicians found the technology to be confusing due to inconsistent automation across different blood work processes.

Conclusions: This study highlights that analytical efficacy alone does not ensure technology use or acceptance. The overall system's design, user perception, and users' knowledge of the technology are equally important and necessary (limitations, functionality, purpose, and scope). Therefore, the effective integration of AI-based decision support systems, such as the BUC, mandates multidisciplinary engagement, ensuring the adequate initial and recurrent training of AI users while maintaining high analytical efficacy and validity. As a final takeaway, the design of AI systems that are made to perform specific tasks must be self-explanatory, so that the users can easily understand how and when to use the technology. Using any technology on a population for whom it was not initially designed will hinder user perception and the technology's use.

(JMIR Hum Factors 2022;9(4):e38411) doi:10.2196/38411

KEYWORDS

artificial intelligence; human factors; decision-making; blood transfusion; technology acceptance; complications; prevention; decision support; transfusion overload; risk; support; perception; safety; usability

Introduction

Blood Transfusion and Challenges

Blood product transfusion (BT) is a critical aspect of routine clinical practice, and over 10.5 million units of blood are transfused annually in hospitals within the United States [1,2]. BT is essential across multiple health care domains [3]. There exists a substantial need for blood, and this need has increased, as the burden of chronic diseases has overlapped with increasing life expectancy [4]. Unfortunately, health care systems have been experiencing BT overuse (unnecessary transfusion), that is, patients are being given more blood than what is physiologically required. The practice of transfusion overuse has been a concern in multiple other countries, including the United Kingdom, Spain, Northern Ireland, and South Africa [5-10]. Transfusion overuse can make patients prone not only to immunological reactions, including hemolysis and acute lung injury, but also to circulatory volume overuse and acute heart failure [11]. In 2011, there were 30 casualties reported among transfusion recipients in the United States, and among all associated risks, transfusion-related acute lung injury and volume overload have been significant causes of morbidity [12,13]. Besides health risks, transfusion overuse also contributes to increased hospital expenses and worsens already limited blood product supplies, resulting in shortages.

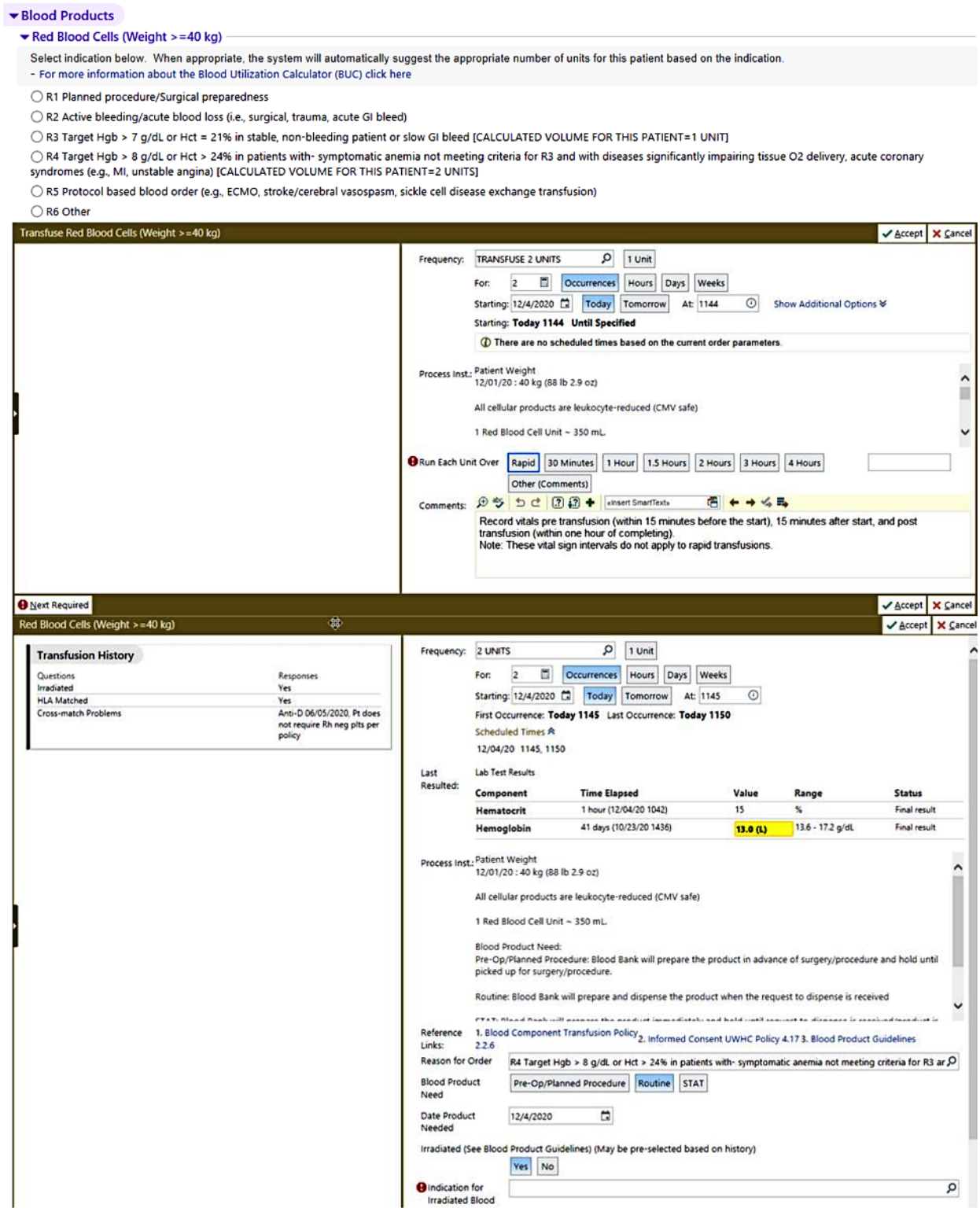
The drawbacks of transfusion overuse have been long identified by authorities and have instigated much interest in institution-based and national patient blood management initiatives within the United States [11,12,14,15]. Additionally, efforts have been entrusted with clinical studies aiming to optimize blood transfusion practices. Research has proposed clinical practice guidelines and processes to standardize blood transfusion. However, noticeable variability in transfusion practices and related outcomes for patients remains. Deciding to transfuse a patient is not always straightforward or linear, and this decision cannot be consistently made based on specific criteria [12,16]. The determinants of standardized blood transfusion encompass several variables, including the clinical scenario, patient risk factors, comorbidities, vital signs, the rate of anemia onset [17], the bleeding rate, and many others. No one numerical laboratory value can be used as a definitive guide for blood transfusion [12,16]. Other factors, such as an insufficient understanding of transfusion guidelines and the diverse recommendations of medical societies, can also contribute to inconsistencies in blood transfusion practices.

Blood Utilization Calculation

To minimize transfusion overuse, a proprietary artificial intelligence (AI)–based blood utilization calculator (BUC) was developed and integrated into the electronic health record at a university hospital in Wisconsin. It is a module of an electronic decision support program known as the Digital Intern (Integrated Vital Medical Dynamics, LLC), and it was designed to ensure the standardization of red blood cell transfusion, following the blood transfusion guidelines. This digital technology runs on a proprietary AI algorithm that provides clinical recommendations for the number of packed red blood cells required to achieve the target hemoglobin or hematocrit value for a given adult patient. It has been reported that the target hemoglobin value was achieved in more than 96% of prescribed transfusions with the help of the BUC (Figure 1) [6,18,19]. It has also been pointed out that the BUC is more consistent than clinicians [18]. Further details of the BUC have been explained elsewhere [20].

Despite its promising performance, the BUC remains underused in the clinical setting. Clinicians often reject BUC recommendations [19], resulting in transfusion overuse and related expenses. Therefore, this qualitative study aims to explore how clinicians perceived this AI-based decision support system and, consequently, understand the factors hindering BUC use.

Figure 1. The artificial intelligence–based blood utilization calculator.



Methods

Ethics Approval

This study took place in collaboration with a university hospital in Wisconsin. It obtained ethical approval from the institutional review board of the Stevens Institute of Technology, Hoboken (institutional review board ID: 2022-021N).

Semistructured Interviews

This study used a qualitative analysis of semistructured interviews to explore the factors affecting clinicians' decision-making regarding blood transfusion. Table 1 shows the interview guide. The fundamental principle of qualitative interviewing is to provide a framework for participants to express their understanding on their terms [21]. Semistructured interviews are typically used in qualitative research and are among the most common data sources in health care research

[22]. They consist of several key questions that not only help define the areas to be explored but also allow interviewers or interviewees to diverge from pursuing an idea or response in more detail [23]. Most importantly, the flexibility of this approach (in comparison to structured or unstructured interviews) enables interviewers to stay focused on their research agenda and allows for the discovery or elaboration of information that is important to participants but may not have previously been thought of as pertinent by the research team [23].

Table 1. Interview guide.

Topic	Guiding questions	Possible follow-up questions
General experience with the BUC[a]	"I am curious to know how you feel when using BUC"	• "How did it impact your clinical performance?" • "How does using BUC impact your decision-making?"
General experience with the BUC	"What are your thoughts about the impact of BUC on patient safety"	• "Under what conditions do you think BUC can cause harm to the patient or give the wrong recommendation?" • "Can you please share your experience when BUC helped you perform better?" • "Did it ever happen when BUC helped you correct or negatively impacted your decision – can you elaborate on that with an example"
Workload	"Clinicians are often overloaded with work. How do you feel BUC has helped reduce or increase some of your workloads?"	• "Can you give an example when BUC made things easier, which otherwise would require more work" • "Can you give an example when BUC made things difficult or confusing that otherwise would be easy"
Decision-making	"When you give the final recommendation about the number of blood units to be transfused for a patient – how do you know when to go with the BUC recommendation and when to make your judgment?"	• "Can you elaborate on how you make a judgment when the BUC recommendation contradicts your decision?" • "Have you ever changed your decision after looking at the BUC recommendation? Can you please elaborate?"
Closure	"Thank you for taking the time to share your BUC experience. Is there anything else you think I should know?"	• N/A[b]

[a]BUC: blood utilization calculator.

[b]N/A: not applicable.

Data Collection and Analysis

We interviewed 10 clinicians (BUC users) until the data saturation point was reached [24]. The saturation sampling method is a well-known methodological principle in qualitative research. It is used to determine, based on the data that have been collected and analyzed, whether further data collection is unnecessary [24]. We decided to stop recruitment after the 10th interview, as we attained thematic saturation. Moreover, a high degree of consensus had begun to emerge among the clinicians who were interviewed, and the information retrieved was sufficient for satisfying the aims of this investigation. According to the literature, data saturation can be reached with 9 to 17 interviews [25].

The interviews were conducted over a web-based platform and were recorded. The audiovisual recordings were then anonymously transcribed verbatim. Each participant was given a US $50 gift card for completing the interview. Each interview lasted approximately 20 to 30 minutes. We used an inductive-deductive thematic analysis to analyze the transcripts [26], which involved applying predetermined themes to the data (deductive) and consecutively identifying new themes as they emerged in the data (inductive) [27]. This method included the interpretation of the text and an analysis of what the text discussed, specifically identifying work system elements and cognitive human factors that influenced clinicians' use of the BUC and clinical decision-making. We also prepared the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Multimedia Appendix 1).

Results

Overview of Clinicians' Perceptions

A total of 10 clinicians from different clinical departments participated in this study. All of the participants were frequent users of the BUC (at least once per week) and had used the technology for at least 1 year. As shown in Textbox 1, we identified the following two themes: (1) workload and usability and (2) clinical decision-making. We discussed each theme briefly by providing detailed quotations.

Textbox 1. Clinicians' perceptions of the artificial intelligence–based blood utilization calculator (BUC).

Workload and usability

- Sample quotes reflecting a negative perception
  - "I remember the first time I saw it, like sort of reading through a lot of options, like, which one of these do apply to me so that I can get the transfusion order to go through, which can take a little bit of time."
  - "It requires quite a few extra clicks to go through the other indication to get it, to let you give plasma for this indication that is often recommended by hepatology."
  - "BUC slowed me down whenever I had to figure out how to bypass the BUC system to get the threshold, we knew we needed for that particular patient."
  - "I think there is a risk of getting down in the cognitive fatigue of decision-making and figuring out which box to click."
- Sample quotes reflecting a positive perception
  - "Earlier, we used to decide how many units like haphazardly, but with BUC, I like that it does part of my thinking. Well, I would say it's easier because now I don't have to think as it tells me how many units of blood, I need to give a patient."
  - "Overall, I like using it because it takes a lot of the thinking out of it in terms of calculations and stuff. BUC makes it a lot easier for every standard patient care."
  - "I think it's pretty easy to use. It's straightforward. It just kind of like leads you exactly through the process."
  - "I find it relatively easy to work with, and I like it because it's straightforward. I just choose whatever I want, and it calculates or puts in the numbers."
  - "I think it's user-friendly and easy. I don't think it adds extra work. I've used it in general surgery trauma, like when you're doing more complicated, like a resuscitation, like one-to-one to one ratio and, um, I think it was pretty user-friendly for that as well."

Clinical decision-making

- Sample quotes reflecting a negative perception
  - "BUC does not improve my decision-making. It's not groundbreaking in any way. It's just more of a reminder of what I have already been taught as a young physician."
- Sample quotes reflecting a positive perception
  - "Overall, it helps you, and I think it helps, you know, determine how much blood a patient need. Kind of making sure we're ordering blood for the right patient indication, reminding us of appropriate criteria. So, I think overall; I think it's a pretty useful tool that we use."
  - "I think it's helpful because it explains like hemoglobin of several patients. If a patient has low platelets, you might have a higher hemoglobin goal. Um, so it's nice to have that spelled out for you, so you don't have to look it up elsewhere and then come back and make the decisions."
  - "I like having the guidelines built-in so that you know when you're doing something that is, um, the, that is the guideline or evidence based. And, you know, when you are deviating from that and therefore hopefully have a good reason for it and are at least cognizant of the fact that you're deviating."
  - "I remember, yeah, a couple of times where we initially wanted to give like two packets of blood, but then [BUC] recommended only one, and we kind of went back and we're like, well, I think the tool is right. Like, we only need to give one unit of blood in this case."
  - "If BUC is telling me that I'm ordering too much blood, I go back, thinking, okay, does the patient need this much blood? So, it's more like I'm ensuring I follow the standard of care, except for those exceptional patient circumstances."

Workload and Usability

For standard care patients, BUC use often helps to standardize blood transfusion and minimize cognitive workload. Overall, clinicians found the BUC to be user-friendly and intuitive. They acknowledged that the BUC has an easy learning curve. Clinicians attending patients with trauma found the BUC user-friendly. However, the perceptions of BUC-related workload were not consistent across all users. According to some clinicians, the BUC was an add-on to their clinical work; they found the BUC interface to be complex for new users. A clinician noted that the interface of the BUC can result in confusion and incorrect transfusion dosages due to inconsistent automation across different blood work processes.

Clinicians' Decision-making and BUC Performance

Clinicians found the BUC to be a helpful technology that often assisted them in making informed clinical decisions regarding blood transfusion, but it did not necessarily improve their decision-making. By providing necessary information regarding transfusion goals, the BUC helped clinicians make faster decisions. They acknowledged the benefits of having the BUC, which enabled them to adhere to the transfusion guidelines. It encouraged them to think critically about their patients and BT practices. Another critical finding was how clinicians made transfusion decisions when their intuition contradicted the BUC. Clinicians said that they consulted their seniors or followed their judgments whenever their decisions failed to match the

BUC recommendations. In other words, clinicians typically trusted the BUC when its recommendations matched their assessments or when a patient had a very standard clinical status, no health complications, or no notable health history. Clinicians also acknowledged that they bypassed BUC recommendations when the recommendations did not match their judgments. Nevertheless, in a few instances, clinicians considered BUC recommendations and changed their judgments after revisiting the patient's health status. In some other cases, the BUC encouraged discussion among clinicians and provided them with an opportunity to adhere to transfusion guidelines.

Discussion

The importance of human factors and AI in health care has been well established by several studies and reputed authorities across all significant health care establishments. This is the first study to explore clinicians' perceptions of an AI-based BUC (ie, an AI decision support system).

Workload and Usability

Clinicians have limited time in their visits and are often overloaded with the burden of clinical documentation. Integrating user-friendly, AI-based decision support systems can effectively assist clinicians and reduce their workloads. Developing a user-friendly and safe technology mandates human factors consideration. Human factors enable us to understand the importance of users' needs and how they may vary based on users' expertise, their environment, and the sensitivity of patients. In our study, depending on their clinical expertise and the patient type, different clinicians perceived the BUC differently. Some found the BUC useful, while others perceived it to be confusing and hard to use, since the technology was not tailored to their needs.

Certain users were not sure when to use and when not to use this technology and oftentimes used the BUC for situations that were beyond its scope (eg, on pediatric patients or patients with sickle cell disease). These users developed a negative perception of the BUC because it was not performing as per their expectations. The BUC is not designed for patients with internal bleeding or sickle cell disease or for ordering blood for scheduled surgeries. It was only built to analyze a given blood value and recommend a transfusion volume to help clinicians achieve their self-selected target blood level. However, trying to use the BUC on other patient types or for other purposes, at times, negatively influenced users' perceptions of the technology. Clinicians often had to figure out a way to bypass the system and place their blood transfusion order, adding to their existing workloads and slowing down the transfusion process. Nevertheless, when the BUC was used on the appropriate patient population, clinicians found it user-friendly and acknowledged that the technology helped reduce their cognitive workloads and, overall, assisted them with their BT tasks and related decision-making.

User-centered design, wherein the user is centrally involved in all phases of the design process, is essential for AI health care technologies. However, designing user-friendly technologies becomes challenging when the user environments and activities are varied (eg, uncrossed transfusion, massive transfusion, etc). This study shows that usability issues can worsen due to the heterogeneity of applications, users' needs, and how users use the technology. The unclear design of AI technologies can result in added workloads; increase the likelihood of patient harm; and, most importantly, hinder clinicians' intent to use the technology. Therefore, adequate training and clarification on the scope, functionality, limitations, and role of a given technology are important for wider acceptance and use.

According to our findings, one way to improve BUC use and acceptance is to have a tailored interface design that automatically detects the treated population based on existing electronic medical record data, the time when a transfusion needs to be ordered (eg, immediately), and the purpose of a transfusion (eg, potential operative need). This approach can ensure that clinicians are shown commonly used information, along with options that are relevant to their patients' needs at a given moment. A tailored BUC design would also ensure selective situation awareness. For example, allowing clinicians to concentrate on relevant details about their patients may help them avoid unnecessary working memory use. Additionally, implementing functions that prevent the BUC from being used on patients who do not fall within its scope can help minimize errors and prevent clinicians from developing a negative perception of the technology. This can be achieved either by incorporating an alert system within the BUC that would flag every time a user uses the technology on any patient outside of the target population or by completely disabling the BUC whenever an incorrect patient type is detected.

Clinical Decision-making

One of this study's main contributions, as well as its novelty, is that it captured the impact of an AI-based decision support system (ie, the BUC) on clinical decision-making. We did not notice any negative impact of the BUC on clinical decision-making. Clinical decision-making is a complex process that necessitates a multidisciplinary systemic approach, encompassing psychology, cognition, and statistics. It is considered a context-driven, time-dependent, and evolving process that requires data collection, interpretation, and evaluation to select the appropriate choice of action [28]. For example, the choice of how much blood should be transfused to a specific patient depends on their body weight, medical status, medical history, rate of blood loss (if any), and treatment plan, among many other factors. Due to such factors, care coordination [29] and shared decision-making in clinical practices are challenging. Our findings indicate the positive impact of the BUC on clinicians' decision-making; the technology acted as an assistive digital platform, promoting well-informed BT. Such impacts of AI have been seen in other fields of medicine [30,31].

In the literature on decision-making, intuitive and analytical decision-making [32] are the two predominant decision-making styles. Intuitive decision-making has been portrayed as an automatic [33] decision process that can be shaped by the work environment and contextual skills [34,35]. Senior clinicians have been observed to prefer the intuitive approach [36]. Their tendency to use the intuitive approach is due to their experience

and ability to make faster and more accurate clinical decisions [37]. Our study captured the same tendency, as attending clinicians seldom considered BUC recommendations. Almost in every situation, when the recommendations generated by the BUC contradicted senior clinicians' judgments, they always followed their judgments, thereby exhibiting confirmation bias.

This study has limitations. It was a single-institution assessment that was conducted within an academic health care establishment. Further, the clinicians who participated were a convenience sample, which introduced self-selection bias. Additionally, the clinicians, per their clinical specialties, were not those who advised or performed blood transfusions the most often. However, a diverse population of clinicians, in terms of clinical expertise, was recruited. Future longitudinal research may help quantify the BUC's impact on patient safety.

Conclusion

This study highlights that analytical efficacy alone does not ensure technology use or acceptance. The overall system's design, user perception, and users' knowledge of the technology are equally important and necessary (limitations, functionality, purpose, and scope). Therefore, the effective integration of AI-based decision support systems, such as the BUC, mandates multidisciplinary engagement, ensuring the adequate initial and recurrent training of AI users while maintaining high analytical efficacy and validity. As seen in this study, all clinicians had different needs that the BUC did not fully address, and the fact that the system's design was not indicative of its actual purpose or target patient population confused its users and hindered its use in the hospital.

As a final takeaway, an AI technology such as the BUC, if not designed for individual users at the department level, might not be used as intended. The design of such AI systems that are made to perform specific tasks must be self-explanatory, so that the users can easily understand how and when to use the technology. AI technologies in health care are only designed and developed to help clinicians identify patterns they would typically overlook. Nevertheless, if clinicians only consider AI recommendations when such recommendations complement their professional and personal judgments or use AI technology on the wrong population, then the motives for having an AI technology in the first place would be in vain.

Conflicts of Interest

The authors declare the following financial interests and personal relationships, which may be considered potential competing interests: JEM invented the Digital Intern artificial intelligence technology that includes the blood utilization calculator, and JEM is an owner and manager of Integrated Vital Medical Dynamics, LLC. The other authors (AC and OA) declare no conflicts of interest.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File , 20 KB - humanfactors_v9i4e38411_app1.docx]

References

1. Sadana D, Kummangal B, Moghekar A, Banerjee K, Kaur S, Balasubramanian S, et al. Adherence to blood product transfusion guidelines-An observational study of the current transfusion practice in a medical intensive care unit. Transfus Med 2021 Aug;31(4):227-235. [doi: 10.1111/tme.12771] [Medline: 33749043]
2. Jones JM, Sapiano MRP, Savinkina AA, Haass KA, Baker ML, Henry RA, et al. Slowing decline in blood collection and transfusion in the United States - 2017. Transfusion 2020 Mar;60 Suppl 2(Suppl 2):S1-S9 [FREE Full text] [doi: 10.1111/trf.15604] [Medline: 32086817]
3. Armstrong B. Benefits and risks of transfusion. ISBT Sci Ser 2008 May 09;3(2):216-230 [FREE Full text] [doi: 10.1111/j.1751-2824.2008.00199.x]
4. Bediako AA, Ofosu-Poku R, Druye AA. Safe blood transfusion practices among nurses in a major referral center in Ghana. Adv Hematol 2021 Mar 02;2021:6739329 [FREE Full text] [doi: 10.1155/2021/6739329] [Medline: 33747086]
5. Adedayo T, O'Mahony D, Adeleke O, Mabunda S. Doctors' practice and attitudes towards red blood cell transfusion at Mthatha Regional Hospital, Eastern Cape, South Africa: A mixed methods study. Afr J Prim Health Care Fam Med 2021 Jun 24;13(1):e1-e8 [FREE Full text] [doi: 10.4102/phcfm.v13i1.2889] [Medline: 34212740]
6. Connor JP, Cunningham AM, Raife T, Rose WN, Medow JE. Standardization of transfusion practice in organ donors using the Digital Intern, an electronic decision support algorithm. Transfusion 2017 Jun;57(6):1369-1375. [doi: 10.1111/trf.14066] [Medline: 28266049]
7. Joy PJ, Bennet SJ. The appropriateness of blood transfusion following primary total hip replacement. Ann R Coll Surg Engl 2012 Apr;94(3):201-203 [FREE Full text] [doi: 10.1308/003588412X13171221501384] [Medline: 22507728]
8. Díaz MQ, Borobia AM, Erce JAG, Maroun-Eid C, Fabra S, Carcas A, USEES-URG Research Group. Appropriate use of red blood cell transfusion in emergency departments: a study in five emergency departments. Blood Transfus 2017 May;15(3):199-206 [FREE Full text] [doi: 10.2450/2016.0324-15] [Medline: 27416566]

9. Barr PJ, Donnelly M, Cardwell CR, Parker M, Morris K, Bailie KEM. The appropriateness of red blood cell use and the extent of overtransfusion: right decision? Right amount? Transfusion 2011 Aug;51(8):1684-1694. [doi: 10.1111/j.1537-2995.2011.03130.x] [Medline: 21470238]
10. Salverda M, Ketharanathan N, Van Dijk M, Beltchev E, Buys H, Numanoglu A, et al. A review of blood transfusions in a trauma unit for young children. S Afr Med J 2017 Feb 27;107(3):227-231. [doi: 10.7196/SAMJ.2017.v107i3.11307] [Medline: 28281428]
11. Mehta N, Murphy MF, Kaplan L, Levinson W. Reducing unnecessary red blood cell transfusion in hospitalised patients. BMJ 2021 Apr 06;373:n830. [doi: 10.1136/bmj.n830] [Medline: 33824140]
12. Goodnough LT, Levy JH, Murphy MF. Concepts of blood transfusion in adults. Lancet 2013 May 25;381(9880):1845-1854. [doi: 10.1016/S0140-6736(13)60650-9] [Medline: 23706801]
13. Vamvakas EC. Reasons for moving toward a patient-centric paradigm of clinical transfusion medicine practice. Transfusion 2013 Apr;53(4):888-901. [doi: 10.1111/j.1537-2995.2012.03825.x] [Medline: 22882177]
14. American Association of Blood Banks: Five things physicians and patients should question. Choosing Wisely. URL: https://www.choosingwisely.org/societies/american-association-of-blood-banks/ [accessed 2022-02-18]
15. Padhi S, Kemmis-Betty S, Rajesh S, Hill J, Murphy MF, Guideline Development Group. Blood transfusion: summary of NICE guidance. BMJ 2015 Nov 18;351:h5832. [doi: 10.1136/bmj.h5832] [Medline: 26581483]
16. Goodnough LT, Audet AM. Utilization review for red cell transfusions. Are we just going through the motions? Arch Pathol Lab Med 1996 Sep;120(9):802-803. [Medline: 9140283]
17. Goodnough LT, Despotis GJ, Hogue CWJ, Ferguson TBJ. On the need for improved transfusion indicators in cardiac surgery. Ann Thorac Surg 1995 Aug;60(2):473-480. [doi: 10.1016/0003-4975(95)98960-3] [Medline: 7646127]
18. Connor JP, Raife T, Medow JE. Outcomes of red blood cell transfusions prescribed in organ donors by the Digital Intern, an electronic decision support algorithm. Transfusion 2018 Feb;58(2):366-371. [doi: 10.1111/trf.14424] [Medline: 29194652]
19. Connor JP, Raife T, Medow JE, Ehlenfeldt BD, Sipsma K. The blood utilization calculator, a target-based electronic decision support algorithm, increases the use of single-unit transfusions in a large academic medical center. Transfusion 2018 Jul;58(7):1689-1696. [doi: 10.1111/trf.14637] [Medline: 29717482]
20. Choudhury A, Asan O, Medow JE. Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system -- Blood Utilization Calculator. Appl Ergon 2022 May;101:103708. [doi: 10.1016/j.apergo.2022.103708] [Medline: 35149301]
21. Patton MQ. Two decades of developments in qualitative inquiry: A personal, experiential perspective. Qual Soc Work 2002 Sep;1(3):261-283. [doi: 10.1177/1473325002001003636]
22. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Fam Med Community Health 2019 Mar 08;7(2):e000057 [FREE Full text] [doi: 10.1136/fmch-2018-000057] [Medline: 32148704]
23. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ 2000 Jan 08;320(7227):114-116 [FREE Full text] [doi: 10.1136/bmj.320.7227.114] [Medline: 10625273]
24. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
25. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
26. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014 Oct 16;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
27. Bingham AJ, Witkowsky P. Deductive and inductive approaches to qualitative data analysis. In: Vanover CF, Mihas PA, Saldana J, editors. Analyzing and Interpreting Qualitative Research: After the Interview. Thousand Oaks, California: SAGE Publications; 2021.
28. Tiffen J, Corbridge SJ, Slimmer L. Enhancing clinical decision making: development of a contiguous definition and conceptual framework. J Prof Nurs 2014;30(5):399-405. [doi: 10.1016/j.profnurs.2014.01.006] [Medline: 25223288]
29. Hepp SL, Suter E, Jackson K, Deutschlander S, Makwarimba E, Jennings J, et al. Using an interprofessional competency framework to examine collaborative practice. J Interprof Care 2015 Mar;29(2):131-137. [doi: 10.3109/13561820.2014.955910] [Medline: 25208088]
30. Hötker AM, Da Mutten R, Tiessen A, Konukoglu E, Donati OF. Improving workflow in prostate MRI: AI-based decision-making on biparametric or multiparametric MRI. Insights Imaging 2021 Aug 09;12(1):112 [FREE Full text] [doi: 10.1186/s13244-021-01058-7] [Medline: 34370164]
31. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes. 2019 May Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-11. [doi: 10.1145/3290605.3300468]
32. Falzer PR, Garman DM. Image Theory's counting rule in clinical decision making: Does it describe how clinicians make patient-specific forecasts? Judgm Decis Mak 2012 May;7(3):268-281 [FREE Full text]

33. Custers EJFM. Medical education and cognitive continuum theory: an alternative perspective on medical problem solving and clinical reasoning. Acad Med 2013 Aug;88(8):1074-1080. [doi: 10.1097/ACM.0b013e31829a3b10] [Medline: 23807108]
34. Lockwood C. Clinical judgement and decision-making in nursing and interprofessional healthcare. Int J Evid Based Healthc 2011;9(1):69-70. [doi: 10.1111/j.1744-1609.2010.00203.x]
35. Parker-Tomlin M, Boschen M, Glendon I, Morrissey S. Factors influencing health practitioners' cognitive processing and decision-making style. J Interprof Care 2019;33(5):546-557. [doi: 10.1080/13561820.2018.1551866] [Medline: 30497308]
36. McLaughlin JE, Cox WC, Williams CR, Shepherd G. Rational and experiential decision-making preferences of third-year student pharmacists. Am J Pharm Educ 2014 Aug 15;78(6):120 [FREE Full text] [doi: 10.5688/ajpe786120] [Medline: 25147392]
37. Jung WH, Kim SN, Lee TY, Jang JH, Choi CH, Kang DH, et al. Exploring the brains of Baduk (Go) experts: gray matter morphometry, resting-state functional connectivity, and graph theoretical analysis. Front Hum Neurosci 2013 Oct 02;7:633 [FREE Full text] [doi: 10.3389/fnhum.2013.00633] [Medline: 24106471]

Abbreviations

AI: artificial intelligence
BT: blood product transfusion
BUC: blood utilization calculator
COREQ: Consolidated Criteria for Reporting Qualitative Research

Edited by A Kushniruk; submitted 31.03.22; peer-reviewed by E Urena, A Chaturvedi, PA Laloë; comments to author 28.06.22; revised version received 17.08.22; accepted 06.10.22; published 31.10.22.

Please cite as:
Choudhury A, Asan O, Medow JE
Clinicians' Perceptions of an Artificial Intelligence–Based Blood Utilization Calculator: Qualitative Exploratory Study
JMIR Hum Factors 2022;9(4):e38411
URL: https://humanfactors.jmir.org/2022/4/e38411
doi:10.2196/38411
PMID:36315238

Original Paper

Usability and Teachability of Continuous Glucose Monitoring Devices in Older Adults and Diabetes Educators: Task Analysis and Ease-of-Use Survey

Simon Psavko[1], MA; Noam Katz[1], MS; Tina Mirchi[1], MS; Courtney R Green[1], PhD

Dexcom, Inc, San Diego, CA, United States

Corresponding Author:
Courtney R Green, PhD
Dexcom, Inc
6340 Sequence Dr
San Diego, CA, 92121-4356
United States
Phone: 1 858 529 4128
Email: courtney.green@dexcom.com

Abstract

Background: Continuous glucose monitoring (CGM) devices continuously sense and relay glucose concentration data from the interstitial fluid to a mobile phone or receiver. Older adults benefit from this continuous monitoring of glucose levels. Proper deployment of the sensing wire is facilitated by a specialized applicator.

Objective: Our aim was to assess a new seventh-generation (G7) CGM device (Dexcom, Inc) for use by adults 65 years of age or older and certified diabetes care and education specialists (CDCESs). Ease of use related to intradermal insertion and mobile app setup will be assessed and compared to the fifth- and sixth-generation systems.

Methods: Formal task analysis was conducted to enumerate the number and complexity of tasks associated with CGM deployment. We recruited 10 older adults with no prior CGM experience and 10 CDCESs to assess ease of use through hands-on insertion and initiation of a G7 system followed by a survey and, for older adults, a system usability scale survey.

Results: About half as many tasks are needed to deploy G7 compared to G6. Older adults and CDCESs reported overall high usability of the G7 CGM device. CDCESs noted G7's easier setup compared to previous generations. The system usability scale score for the CGM system was 92.8, which reflects excellent usability.

Conclusions: For CDCESs and for older adults using the G7 CGM system, cognitive burden is relatively low and reduced compared to previous CGM systems. Easing of this burden and simplification of the glucose monitoring aspect of proper diabetes management will likely contribute to improved outcomes in this population.

(JMIR Hum Factors 2022;9(4):e42057) doi:10.2196/42057

KEYWORDS

medical devices; wearable devices; older adults; task analysis; usability testing; continuous glucose monitoring; glucose monitoring; glucose levels; diabetes; usability

Introduction

A substantial proportion of people with diabetes do not reach the goals of treatment [1], which is due in part to barriers to effective self-care. Real-time continuous glucose monitoring (RT-CGM) is recommended for patients with diabetes using insulin [2], and many studies support the utility of continuous glucose monitoring (CGM) devices with respect to reducing HbA1c and improving outcomes [3,4]. Ease of use (objectively or subjectively determined) is critically important because this is a key factor in CGM compliance and favorable clinical outcomes.

Certified diabetes care and education specialists (CDCESs) are health care professionals with expertise in diabetes prevention, prediabetes, and diabetes management. They often work as part of a care team to help those with diabetes understand their disease and achieve their goals in managing diabetes. A recent study found that glycemic and metabolic outcomes were similar in patients who received a telehealth-based diabetes consultation with an endocrinologist or a diabetes self-management education

visit with a CDCES [5]. Because CDCESs train patients who are new to CGM, their comfort with CGM device insertion and app setup is also critical to patient training and adoption.

Management of diabetes has evolved dramatically [6], and CGM use is rapidly proliferating. Still, diabetes self-management imposes cognitive demands and requires a significant level of expertise [7]. Recent studies demonstrated an improvement in glycemic outcomes with CGM for patients 65 years or older [8] that was sustained for an additional 12 months [9]. Although usage in older adults was very high in these studies (at least 83% of participants wore a RT-CGM device ≥6 days/week over the 6-month study period [8]), barriers still exist to broad population-level use. Wildenbos et al [10] categorized these barriers as related to cognition, motivation, physical ability, and perception. This can include specific barriers such as memory decline, arthritis (especially in the hands), impaired vision or hearing, and perception of the technology as too complex [10,11]. These barriers can be lowered through enhanced usability of CGM. RT-CGM has the potential to improve quality of life in older adults, but adoption barriers must be considered and addressed [11].

A seventh-generation CGM system (G7; Dexcom, Inc) received CE mark in March 2022 and provides accurate estimates of glucose levels in the interstitial fluid in adults [12] as well as children and adolescents [13]. Clinical benefits are anticipated due to the similarity of accuracy metrics between the fifth-, sixth-, and seventh-generation systems and an increased feature set [14].

To maximize usability, G7 was designed using human factors engineering and usability engineering processes. Human factors analysis throughout product development has been shown to enhance usability of CGM sensor applicators [15]. In designing the device, participants representative of the intended user population were evaluated on their ability to use the CGM system safely and effectively. This included 3 human factors experts who performed multiple heuristic reviews of the device and 381 participants representative of the intended user population who participated across 26 usability studies, evaluating all aspects of the Dexcom G7 system hardware, software, and labeling. All usability studies were conducted in accordance with US Food and Drug Administration and international guidance [16-19].

There is a paucity of CGM usability studies specifically in special populations such as older adults and diabetes educators. In this study, we describe a task analysis and ease-of-use study conducted with a commercially representative G7 system with 10 CDCESs and 10 CGM-naïve older adults with type 2 diabetes (T2D). Relevant surveys were conducted of each group, and responses to open-ended discussions were recorded.

Methods

Task Analysis

A task analysis and heuristic evaluation was conducted on the fifth-, sixth-, and seventh-generation CGM systems. The task analysis identified the perceptual inputs, cognitive processes, and actions required for a user to complete the task of sensor insertion. The heuristic evaluation was based on existing design principles and compliance with recognized usability guidelines to uncover potential use errors and end user risks. The number of tasks required for sensor insertion was enumerated as well as the potential unrecoverable use errors, which are defined as incorrect actions committed by the user during the sensor insertion process that would result in predeployment loss of the device. Potential errors that would lead to sensor loss include failing to remove one or both of the adhesive backings from the applicator prior to deploying the sensor and sensor deployment before proper placement of the device on the skin.

Usability Study

Although it is anticipated that CGM-naïve patients initiating CGM use would receive one-on-one training, discuss personal-use CGM**,** and be directed to additional web-based training prior to using CGM, this study chose to evaluate a ‘no-training’ scenario for the following 3 reasons: (1) it is difficult to provide an equivalent level of detail across different instances of one-on-one training; (2) one-on-one training may not occur in every real-world instance; and (3) lack of training represents the most stringent evaluation of safe and effective usability.

In an in-person, one-on-one setting, 10 CDCESs and 10 CGM-naïve older adults with T2D were provided with a G7 system and a mobile phone app. Both groups were tasked with setting up the mobile app, deploying the sensor on themselves, and establishing communication between the app and the sensor without training or instructions beyond what is included in the G7 box and mobile app. This included allowing all notifications, pairing the wearable, reviewing safety information, reading the alert functionality explanation, watching the required videos, and following sensor insertion instructions. During the study session, all relevant activities performed by participants, including successful and unsuccessful task completion, root causes, and salient participant questions or comments, were recorded. Their total sensor insertion time was also recorded. Following the initiation of G7, both groups completed a posttest survey. Responses to open-ended questions were also recorded. The older adult cohort also completed a system usability scale (SUS) survey [20] to assess their comfort with CGM hardware and app setup as CGM-naïve users. Both surveys (ie, posttest and SUS) evaluated responses using a 5-point Likert scale (1=strongly disagree and 5=strongly agree).

Ethical Considerations

The G7 CGM device had CE marking at the time of the study. This ease-of-use study was not part of an intervention or trial, and the device was not used to make diagnostic or treatment decisions. Similar to the submission of human factors reports to the Food and Drug Administration, which does not require an institutional review board or independent ethics committee approval [16], these approvals were not sought for this ease-of-use study.

Results

Task Analysis Found Fewer Tasks Required for G7 Insertion

Insertion and initialization of the G7 system was subjected to a task analysis. The number of user tasks required for sensor insertion decreased with each subsequent generation (G5: 17 tasks; G6: 13 tasks; and G7: 6 tasks; Figure 1). In addition, the number of potential unrecoverable use errors also decreased (G5: 8; G6: 5; and G7: 1; Figure 1). The full task analyses of each device are summarized in Tables S1-S3 in Multimedia Appendix 1.

Two groups were recruited to participate in this study. A total of 10 CDCESs with an average of 15.7 years of experience agreed to participate. On average, CDCESs train 18 patients on CGM use per month, and a majority of their patients use CGM (Table 1). The second group consisted of 10 older adults with T2D. Their average age was 69.7 years, all were CGM-naïve, and all were on a multiple daily injection insulin regimen (Table 1).

Figure 1. Summary of tasks required for each generation (fifth generation: G5; sixth generation: G6; seventh generation: G7) of Dexcom's continuous glucose monitoring (CGM) devices.

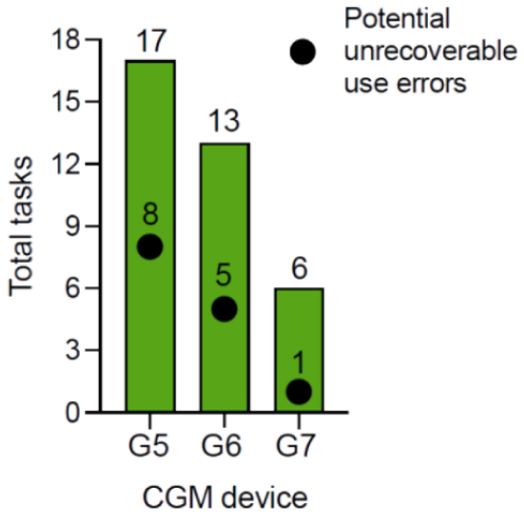


Table 1. Demographics of certified diabetes care and education specialists and older adult participants.

Participants	Values
Certified diabetes care and education specialists (N=10)	
Years of experience, mean (range)	15.7 (5-25)
Patients trained on CGM[a] per month, mean (range)	18 (5-50)
Patients using CGM (%), mean	58.5
Older adults with type 2 diabetes (N=10)	
Age, mean (range)	69.7 (65-79)
CGM-naïve, n (%)	10 (100)
Treatment by multiple daily insulin injections, n (%)	10 (100)

[a]CGM: continuous glucose monitoring.

CDCESs Usability Ratings

All CDCES participants successfully completed G7 app setup and sensor insertion. The time on task was recorded from the start of app onboarding setup through completion of the onboarding (including sensor insertion). The total setup time ranged from 6 minutes to 9.2 minutes (average=7.15 minutes). In the CDCES posttest survey, average ratings were very favorable, and only one participant provided a rating that was assessed to be less than neutral (Table 2). In particular, the statements “the system was easy to learn to use,” “I believe this system is easy to train patients to use,” and “I believe this system is easy to set up” received mean ratings of 5.0 (“strongly agree”) out of a possible score of 5.

In the open-ended discussion, CDCESs were asked to discuss their experience with G7 compared to other CGM devices they use. The participants expressed that setting up and initiating the seventh-generation CGM is extremely easy, requires fewer steps, and feels much less confusing compared to other products. CDCES participants were also asked to discuss their views of virtual CGM training. These participants viewed virtual training of the CGM system as practical and believed it would be very easy for patients to do the training at home on their own.

Table 2. Posttest survey results from interviews with 10 certified diabetes care and education specialists on use of the seventh-generation (G7) continuous glucose monitoring system.

Statement	Mean	Range	Rating and rationale for ratings lower than neutral
I could effectively complete the tasks that were given to me.	4.9	4-5	N/A[a]
I felt comfortable using this system.	4.9	4-5	N/A
The system was easy to learn to use.	5.0	5	N/A
Whenever I made a mistake, I could recover easily and quickly.	4.7	3-5	Rating of 3: participants did not make any mistakes and thus rated this statement as neutral.
The functions I saw worked as I would expect.	4.6	2-5	Rating of 2: participant stated the sound of the sensor insertion deployment was very loud.
The system showed information clearly and effectively.	4.8	4-5	N/A
I found this system unnecessarily complex.	1.0	1	N/A
I believe this system is easy to train patients to use.	5.0	5	N/A
I believe this system is easy to set up.	5.0	5	N/A
I believe a patient can self-train on this system.	4.8	4-5	N/A
It is easier to train on G7 than G6.	4.9	4-5	N/A
G7 setup requires less time to train or set up than G6.	4.9	4-5	N/A

[a]N/A: not applicable.

Older Adults With T2D Usability Ratings

All CGM-naïve older adult participants successfully completed G7 app setup and sensor insertion. The time on task was recorded from the start of the app setup through onboarding completion (including sensor insertion), with setup time ranging from 9 minutes to 18 minutes (average=12.6 minutes). Sensor insertion time was also recorded, with insertion time ranging from 58 seconds to 3 minutes (average=1.95 minutes).

A posttest survey and a SUS survey were given to the older adult participants (Table 3 and Table S4 in Multimedia Appendix 2). No responses lower than neutral were recorded in the posttest survey (Table 3). The SUS score for setup and insertion of the G7 system was 92.8 (Table S4 in Multimedia Appendix 2). This score reflects an excellent usability rating [21].

In the open-ended discussion, older adults were also asked to discuss how comfortable they would be setting up G7 on their own with no training or assistance. All participants stated they would feel “comfortable” or “very comfortable” and specified that the app setup and insertion instructions were concise and easy to follow. They stated that the instructions, images, and in-app videos during setup are helpful. Overall, the G7 system provides increased usability, when compared to previous generations, in terms of efficacy, efficiency, ease of user learning, and user satisfaction.

Table 3. Posttest survey results from interviews with 10 older adults with type 2 diabetes on use of the seventh-generation (G7) continuous glucose monitoring system.

Statement	Mean	Range	Rating and rationale for ratings lower than neutral
I could effectively complete the tasks that were given to me.	4.9	4-5	N/A[a]
I felt comfortable using the G7 system.	4.8	4-5	N/A
The G7 system was easy to learn to use.	4.7	4-5	N/A
The functions I saw worked as I would expect.	4.7	3-5	N/A
The G7 system showed information clearly and effectively.	4.9	4-5	N/A
I believe the G7 system is easy to set up.	4.6	3-5	N/A
I believe I can set up the G7 system on my own.	4.9	4-5	N/A
I believe I can set up the G7 system in a virtual training.	4.8	4-5	N/A

[a]N/A: not applicable.

Discussion

The analyses presented in this paper constitute the first ease of use study of the Dexcom G7 RT-CGM. Previously published results of a task analysis [14] and various survey results presented in this study show that the G7 sensor is easier to insert and set up compared to the two previous device generations. The reduction in potential unrecoverable use errors also reduces

the likelihood of wasting a sensor if a mistake is made during the setup or insertion process. Simpler sensor insertion and app setup processes allow the seventh-generation CGM system to be even easier for older adults to learn and use, which may aid in their motivation to try a new technology [11]. RT-CGM with its alerts and alarms has been shown to contribute to better glycemic outcomes in older adults, including severe hypoglycemia (SH), which is particularly dangerous in this population [8,9]. In a study by the Wireless Innovation for Seniors With Diabetes Mellitus study group [8], CGM users achieved an adjusted difference of –1.9 percentage points in time below 70mg/dL compared to the standard blood glucose monitoring group. This benefit was sustained after 12 months of CGM use [9].

Extensive design work and human factors analyses contributed to the design of the seventh-generation system, and a multifaceted ease-of-use study was described in this paper. This study included an objective task analysis and surveys of CDCESs and older adults with diabetes, which demonstrated robust improvements in the ease of use and overall comfort with the device. A similar study assessed the ease-of-use ratings of the Dexcom G6's automatic sensor applicator versus the Dexcom G5's manual sensor applicator [15]. In our study, we interviewed both older adults with T2D and CDCESs because although people with diabetes are the end users of the device, CDCESs are the primary individuals who would train and educate older adults, and therefore, their comfort with device operation is critical.

CDCESs work on a diabetes care team to develop strategies with patients for successful diabetes management. They play a unique role in helping patients understand their diabetes and learn new technologies [22,23]. Highly CGM-literate CDCESs could aid in CGM adoption by older adults. A recent study of patients with type 1 diabetes covered by Medicare found that CGM adoption remains low, and disparities persist between racial groups [24]. Improved patient education and exposure could reduce these disparities. In a recent study of 171 users of a blood glucose monitoring device with T2D, 29/171 (21%) expressed a lack of interest in CGM adoption [25]; 4/29 (14%) respondents cited a lack of familiarity with CGM as the reason for their disinterest [25]. Recently, more patients gained access to CGM when the Centers for Medicare and Medicaid Services expanded coverage to any CGM that connects to an insulin pump or a standalone receiver and eliminated the requirement for 4 finger sticks per day [26,27]. A proposed local coverage determination from Centers for Medicare and Medicaid Services also would expand CGM coverage to those with a history of problematic hypoglycemia, regardless of insulin use [28]. This is encouraging because the association between CGM use and improved outcomes in people with T2D is strong and growing [3,29,30].

Adoption of CGM in older adults is particularly important for several reasons. The most important aspect of glucose management in older adults is the avoidance of SH. Older patients are more likely to have hypoglycemia unawareness, a reduced ability to produce counterregulatory hormones, and an altered metabolism that increases the risk of SH due to polypharmacy [11]. The risks associated with SH include falls leading to fractures or other injuries, cardiovascular complications, and temporary or permanent cognitive impairment [11,31,32]. Because of the high risk associated with SH in older adults, an International Consensus on Time in Range recommended a more relaxed set of CGM-based targets with the goal of avoiding hypoglycemia [33]. If needed, coordination with a CDCES or other health care providers to adjust alert thresholds and settings could improve user experience.

This study contained some limitations. The small sample size, inclusion of only US participants, and inclusion of only CDCESs (ie, a highly skilled but small group of health care providers) limit the generalizability of the results. This study also did not assess long-term use of the G7 and, because no data were generated, usability of the app interface was not tested.

Ease of use is a critically important element in all diabetes technology, including CGMs. This seventh-generation device was designed using human factors engineering and usability engineering processes, and iterations were performed after reviews from hundreds of potential users. Increased adoption of these technologies relies on appropriate usability, safety, effectiveness, and robust design [34,35]. Older adults report excellent G7 usability and CDCESs report that the G7 is simpler and easier to use compared to the two previous generations of the system. Data presented in this study should inform prescribing and treatment decisions in this vulnerable population.

Acknowledgments

SP, TM, and NK devised the project and performed the task analyses and usability studies. SP, TM, and NK analyzed the data. SP wrote the first draft. CRG wrote the manuscript in consultation with SP, TM, and NK. The authors would like to thank Dr David Price for his role in the study conceptualization and Dr John Welsh for his editorial assistance. This analysis was funded by Dexcom, Inc. Dexcom is a registered trademark of Dexcom, Inc in the United States and other countries.

Conflicts of Interest

The authors are employees of Dexcom, Inc.

Multimedia Appendix 1
Full task analyses of the fifth-, sixth-, and seventh-generation real-time continuous glucose monitoring devices.
[DOCX File , 21 KB - humanfactors_v9i4e42057_app1.docx]

Multimedia Appendix 2
Posttest system usability scale (SUS) results from interviews with 10 older adult participants.
[DOCX File , 15 KB - humanfactors_v9i4e42057_app2.docx]

References

1. Chen Y, Rolka D, Xie H, Saydah S. Imputed state-level prevalence of achieving goals to prevent complications of diabetes in adults with self-reported diabetes - United States, 2017-2018. MMWR Morb Mortal Wkly Rep 2020 Nov 13;69(45):1665-1670 [FREE Full text] [doi: 10.15585/mmwr.mm6945a1] [Medline: 33180755]
2. American Diabetes Association Professional Practice Committee, Draznin B, Aroda VR, Bakris G, Benson G, Brown FM, et al. 7. Diabetes Technology: Standards of Medical Care in Diabetes-2022. Diabetes Care 2022 Jan 01;45(Supplement_1):S97-S112. [doi: 10.2337/dc22-S007] [Medline: 34964871]
3. Beck RW, Riddlesworth TD, Ruedy K, Ahmann A, Haller S, Kruger D, DIAMOND Study Group. Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections: a randomized trial. Ann Intern Med 2017 Sep 19;167(6):365-374. [doi: 10.7326/M16-2855] [Medline: 28828487]
4. Ruedy KJ, Parkin CG, Riddlesworth TD, Graham C, DIAMOND Study Group. Continuous glucose monitoring in older adults with type 1 and type 2 diabetes using multiple daily injections of insulin: results from the DIAMOND trial. J Diabetes Sci Technol 2017 Nov;11(6):1138-1146 [FREE Full text] [doi: 10.1177/1932296817704445] [Medline: 28449590]
5. Nyenwe EA, Ashby S, Tidwell J. Diabetes consultation versus diabetes education in patients with poor glycaemic control: a telemedicine intervention study. J Telemed Telecare 2020 Sep 29;28(9):687-693. [doi: 10.1177/1357633X20959213] [Medline: 32990153]
6. Hirsch IB. Introduction: History of Glucose Monitoring. In: Role of Continuous Glucose Monitoring in Diabetes Treatment. Arlington (VA): American Diabetes Association; Aug 2018.
7. Lippa KD, Klein HA, Shalin VL. Everyday expertise: cognitive demands in diabetes self-management. Hum Factors 2008 Feb;50(1):112-120. [doi: 10.1518/001872008X250601] [Medline: 18354975]
8. Pratley RE, Kanapka LG, Rickels MR, Ahmann A, Aleppo G, Beck R, Wireless Innovation for Seniors With Diabetes Mellitus (WISDM) Study Group. Effect of continuous glucose monitoring on hypoglycemia in older adults with type 1 diabetes: a randomized clinical trial. JAMA 2020 Jun 16;323(23):2397-2406 [FREE Full text] [doi: 10.1001/jama.2020.6928] [Medline: 32543682]
9. Miller KM, Kanapka LG, Rickels MR, Ahmann AJ, Aleppo G, Ang L, et al. Benefit of continuous glucose monitoring in reducing hypoglycemia is sustained through 12 months of use among older adults with type 1 diabetes. Diabetes Technol Ther 2022 Jun;24(6):424-434. [doi: 10.1089/dia.2021.0503] [Medline: 35294272]
10. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
11. Allen NA, Litchman ML. Chapter 10 - Using diabetes technology in older adults. In: Diabetes Digital Health. Cambridge, UK: Elsevier; 2020:131-143.
12. Garg SK, Kipnes M, Castorino K, Bailey TS, Akturk HK, Welsh JB, et al. Accuracy and safety of Dexcom G7 continuous glucose monitoring in adults with diabetes. Diabetes Technol Ther 2022 Jun;24(6):373-380 [FREE Full text] [doi: 10.1089/dia.2022.0011] [Medline: 35157505]
13. Laffel LM, Bailey TS, Christiansen MP, Reid JL, Beck SE. Accuracy of a seventh-generation continuous glucose monitoring system in children and adolescents with type 1 diabetes. J Diabetes Sci Technol 2022 Apr 25:19322968221091816 [FREE Full text] [doi: 10.1177/19322968221091816] [Medline: 35466707]
14. Welsh JB, Psavko S, Zhang X, Gao P, Balo AK. Comparisons of fifth-, sixth-, and seventh-generation continuous glucose monitoring systems. J Diabetes Sci Technol 2022 Jun 13:19322968221099879 [FREE Full text] [doi: 10.1177/19322968221099879] [Medline: 35695305]
15. North R, Pospisil C, Clukey RJ, Parkin CG. Impact of human factors testing on medical device design: validation of an automated CGM sensor applicator. J Diabetes Sci Technol 2019 Sep;13(5):949-953 [FREE Full text] [doi: 10.1177/1932296819831071] [Medline: 30762437]
16. Applying human factors and usability engineering to medical devices. U.S. Food and Drug Administration. 2016 Feb. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices [accessed 2022-11-29]
17. Association for the Advancement of Medical Instrumentation. Human factors engineering - design of medical devices. In: American National Standard. Arlington, VA: Association for the Advancement of Medical Instrumentation; 2013.
18. U.S. Department of Health and Human Services Food and Drug Administration. Design considerations for devices intended for home use. Guidance for Industry and Food and Drug Administration Staff. 2014 Nov 24. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/design-considerations-devices-intended-home-use [accessed 2022-11-29]
19. IEC 62366-1:2015(en) Medical devices—Part 1: application of usability engineering to medical devices. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iec:62366:-1:ed-1:v1:en [accessed 2022-11-28]
20. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.

21. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J of User Experience 2009 May;4(3):114-123 [FREE Full text]
22. Greenwood DA, Howell F, Scher L, Yousef G, Rinker J, Yehl K, et al. A framework for optimizing technology-enabled diabetes and cardiometabolic care and education: the role of the diabetes care and education specialist. Diabetes Educ 2020 Aug 11;46(4):315-322. [doi: 10.1177/0145721720935125] [Medline: 32780001]
23. MacLeod J, Scher L, Greenwood D, Isaacs D, Albanese-O'Neill A, O'Neill MS, et al. Technology disparities and therapeutic inertia: a call to action for the diabetes care and education specialist. ADCES in Practice 2021 Sep;9(5):34-41. [doi: 10.1177/2633559x211032227]
24. Wherry K, Zhu C, Vigersky RA. Inequity in adoption of advanced diabetes technologies among Medicare fee-for-service beneficiaries. J Clin Endocrinol Metab 2022 Apr 19;107(5):e2177-e2185. [doi: 10.1210/clinem/dgab869] [Medline: 34910144]
25. Sears L, Imrisek SD, Hoy-Rosas JR, Lavaysse LM, Lee M, Chapman MR, et al. 705-P: CGM attitudes and adoption among people with type 2 diabetes using One Drop. Diabetes 2022 Jun 01;71(Supplement_1):705-P. [doi: 10.2337/db22-705-p]
26. On Medicare? Accessing CGMs just got easier. Juvenile Diabetes Research Foundation. URL: https://www.jdrf.org/blog/2021/07/23/on-medicare-accessing-cgms-just-got-easier/ [accessed 2022-10-10]
27. More CGMs now covered by Medicare. Juvenile Diabetes Research Foundation. 2021 Jul 23. URL: https://www.jdrf.org/blog/2022/03/10/more-cgms-now-covered-by-medicare/ [accessed 2022-10-10]
28. DL33822: Glucose Monitors. Centers for Medicare and Medicaid Services. 2022 Oct 06. URL: https://www.cms.gov/medicare-coverage-database/view/lcd.aspx?lcdid=39473&ver=16&bc=0 [accessed 2022-10-11]
29. Haak T, Hanaire H, Ajjan R, Hermanns N, Riveline J, Rayman G. Flash glucose-sensing technology as a replacement for blood glucose monitoring for the management of insulin-treated type 2 diabetes: a multicenter, open-label randomized controlled trial. Diabetes Ther 2017 Feb;8(1):55-73 [FREE Full text] [doi: 10.1007/s13300-016-0223-6] [Medline: 28000140]
30. Wright EE, Subramanian S. Evolving use of continuous glucose monitoring beyond intensive insulin treatment. Diabetes Technol Ther 2021 Sep;23(S3):S12-S18. [doi: 10.1089/dia.2021.0191] [Medline: 34546082]
31. Chow L, Seaquist ER. How significant is severe hypoglycemia in older adults with diabetes? Diabetes Care 2020 Mar;43(3):512-514 [FREE Full text] [doi: 10.2337/dci19-0069] [Medline: 32079685]
32. Lipska KJ, Ross JS, Wang Y, Inzucchi SE, Minges K, Karter AJ, et al. National trends in US hospital admissions for hyperglycemia and hypoglycemia among Medicare beneficiaries, 1999 to 2011. JAMA Intern Med 2014 Jul;174(7):1116-1124 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1824] [Medline: 24838229]
33. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019 Aug;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
34. Klonoff DC, Kerr D. Overcoming barriers to adoption of digital health tools for diabetes. J Diabetes Sci Technol 2018 Jan;12(1):3-6 [FREE Full text] [doi: 10.1177/1932296817732459] [Medline: 28969460]
35. Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. Diabetes Technol Ther 2016 Feb;18 Suppl 2(Suppl 2):S3-S13 [FREE Full text] [doi: 10.1089/dia.2015.0417] [Medline: 26784127]

Abbreviations

CDCES: certified diabetes care and education specialist
CGM: continuous glucose monitoring
RT-CGM: real-time continuous glucose monitoring
SH: severe hypoglycemia
SUS: System Usability Scale
T2D: type 2 diabetes

Edited by A Kushniruk; submitted 19.08.22; peer-reviewed by N Allen, S Schembre; comments to author 04.10.22; revised version received 18.10.22; accepted 08.11.22; published 15.12.22.

Please cite as:
Psavko S, Katz N, Mirchi T, Green CR
Usability and Teachability of Continuous Glucose Monitoring Devices in Older Adults and Diabetes Educators: Task Analysis and Ease-of-Use Survey
JMIR Hum Factors 2022;9(4):e42057
URL: https://humanfactors.jmir.org/2022/4/e42057
doi:10.2196/42057
PMID:36347498

Original Paper

Acceptability and Feasibility of a Low-Cost Device for Gestational Age Assessment in a Low-Resource Setting: Qualitative Study

Angela Koech[1,2], MBChB, MSc, MMed; Peris Muoga Musitia[1,3], BSc, MA; Grace Mkanjala Mwashigadi[1], BSc, MPH; Mai-Lei Woo Kinshella[4], BA, MA; Marianne Vidler[4], BA, MPH, PhD; Marleen Temmerman[1,2], BA, MMed, MPH, PhD; Rachel Craik[5,6], BSc; Peter von Dadelszen[6], DipObst, BMedSci, MBChB, DPhil; J Alison Noble[7], BSc, MSc, DPhil; Aris T Papageorghiou[5,8], MBChB, PhD; The PRECISE Network[9]

[1]Centre of Excellence in Women & Child Health, Aga Khan University, Nairobi, Kenya
[2]Department of Obstetrics and Gynaecology, Aga Khan University, Nairobi, Kenya
[3]Health Services Unit, Kenya Medical Research Institute Wellcome Trust Research Programme Nairobi, Nairobi, Kenya
[4]Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada
[5]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, United Kingdom
[6]Department of Women and Children's Health, King's College London, London, United Kingdom
[7]Department of Engineering Science, University of Oxford, Oxford, United Kingdom
[8]Oxford Maternal & Perinatal Health Institute, University of Oxford, Oxford, United Kingdom
[9]See Authors' Contributions

Corresponding Author:
Angela Koech, MBChB, MSc, MMed
Centre of Excellence in Women & Child Health
Aga Khan University
3rd Parklands Avenue
Limuru Road, P.O. Box 30270-00100
Nairobi, 4100
Kenya
Phone: 254 722 502602
Email: angela.koech@aku.edu

Abstract

Background: Ultrasound for gestational age (GA) assessment is not routinely available in resource-constrained settings, particularly in rural and remote locations. The TraCer device combines a handheld wireless ultrasound probe and a tablet with artificial intelligence (AI)-enabled software that obtains GA from videos of the fetal head by automated measurements of the fetal transcerebellar diameter and head circumference.

Objective: The aim of this study was to assess the perceptions of pregnant women, their families, and health care workers regarding the feasibility and acceptability of the TraCer device in an appropriate setting.

Methods: A descriptive study using qualitative methods was conducted in two public health facilities in Kilifi county in coastal Kenya prior to introduction of the new technology. Study participants were shown a video role-play of the use of TraCer at a typical antenatal clinic visit. Data were collected through 6 focus group discussions (N=52) and 18 in-depth interviews.

Results: Overall, TraCer was found to be highly acceptable to women, their families, and health care workers, and its implementation at health care facilities was considered to be feasible. Its introduction was predicted to reduce anxiety regarding fetal well-being, increase antenatal care attendance, increase confidence by women in their care providers, as well as save time and cost by reducing unnecessary referrals. TraCer was felt to increase the self-image of health care workers and reduce time spent providing antenatal care. Some participants expressed hesitancy toward the new technology, indicating the need to test its performance over time before full acceptance by some users. The preferred cadre of health care professionals to use the device were antenatal clinic nurses. Important implementation considerations included adequate staff training and the need to ensure sustainability and consistency of the service. Misconceptions were common, with a tendency to overestimate the diagnostic capability, and expectations that it would provide complete reassurance of fetal and maternal well-being and not primarily the GA.

Conclusions: This study shows a positive attitude toward TraCer and highlights the potential role of this innovation that uses AI-enabled automation to assess GA. Clarity of messaging about the tool and its role in pregnancy is essential to address misconceptions and prevent misuse. Further research on clinical validation and related usability and safety evaluations are recommended.

(JMIR Hum Factors 2022;9(4):e34823) doi:10.2196/34823

KEYWORDS

gestational age; gestation; gynecology; gynecologist; prenatal; antenatal; maternal; fetus; fetal; ultrasound; imaging; pregnancy dating; handheld; portable; trust; artificial intelligence; sub-Saharan Africa; Africa; low cost; LMIC; low income; feasibility; acceptability; AI; pregnancy; pregnant; maternity; women's health; obstetrics; obstetrician; rural; remote; remote location; misconception; eHealth; digital health

Introduction

Knowledge of gestational age (GA) informs decisions in maternal and neonatal care [1], such as the use of corticosteroids in suspected preterm labor [2] and timing of delivery in postterm pregnancy and other pregnancy complications [3]. Reliable estimation of GA improves care by guiding these decisions and reducing unnecessary interventions. It also enables more accurate categorization of low-birth-weight babies into preterm or small for gestational age [4], impacting care for these babies [5,6] and improving reporting of perinatal outcomes [1,7,8].

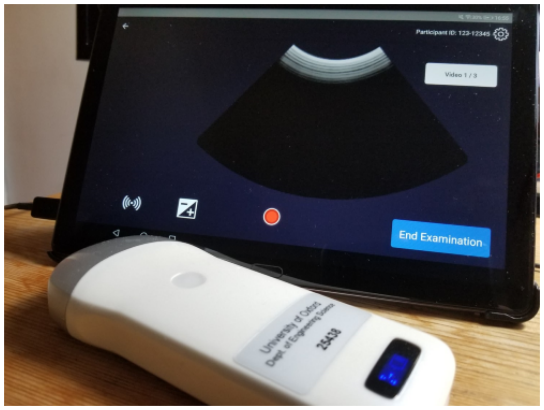
In most high-income country settings, accurate pregnancy dating is provided routinely through early pregnancy ultrasound [9]. However, routine pregnancy ultrasound is rarely available in low- and middle-income country (LMIC) settings. In addition, other methods of GA assessment are likely to be less reliable: recall of last menstrual period (LMP) is generally poor [10], there is frequent late initiation of antenatal care (ANC) [7,11], and there is a higher prevalence of fetal growth restriction [12].

Setting up sustainable routine pregnancy ultrasound services in resource-constrained LMIC settings is often difficult, particularly in rural and remote locations [13]. ANC is largely provided by nurses and nurse-midwives and understaffing is frequent [14]. Radiologists, sonographers, and obstetricians are limited, located primarily in urban areas, and burdened with managing pregnancy complications and clinical emergencies [13]. Skills for GA assessment by ultrasound often require lengthy training programs, regular quality control, and close supervision, all of which are difficult to achieve and sustain [15]. In addition, conventional ultrasound equipment is expensive and requires regular maintenance and appropriate infrastructure such as a reliable continuous power supply.

The TraCer GA assessment device automates the measurement of the fetal transcerebellar diameter (TCD) and head circumference (HC). TraCer uses a low-cost, commercially available handheld battery-powered wireless ultrasound probe, which is linked, via Wi-Fi, to software running on a consumer-grade Android tablet (Figure 1). The device guides and assists health care workers (HCWs) to obtain ultrasound videos of the fetal head. GA is then estimated from the TCD and HC using semiautomated (and, in the future, automated) image recognition and analysis. The TCD estimates cerebellar size, which is considered a good measure for GA, as it is predictable throughout pregnancy and is not heavily impacted by the existence of fetal growth restriction [16,17]. The method also uses the fetal HC, because cerebellar imaging at advanced gestation stages may not always be possible using a low-cost device.

TraCer has been designed to specifically address challenges in resource-limited LMIC settings. However, this does not guarantee that it would be implementable or that communities and health care providers would find it acceptable [18]. A review of innovative approaches for improving maternal and newborn health found that gaps in understanding feasibility, appropriateness, and acceptability of implementation can compromise their capability for effective scale-up [19]. Therefore, it is recommended that new tools and innovations be evaluated not only for their technical and clinical performance, but also for acceptability and appropriateness, usability, and the feasibility of implementation within the intended settings [19].

Figure 1. TraCer Device: the wireless handheld probe and TraCer software on the tablet.



Toward this end, in this study, we assessed the perceptions of pregnant women, their family members, HCWs, and managers regarding the acceptability and feasibility of the TraCer device in health facilities. This will guide further device development and inform plans for clinical implementation.

Methods

Study Design and Setting

A cross-sectional, descriptive, qualitative study was conducted in two public health facilities in Kilifi county in coastal Kenya. The two health facilities are the rural Rabai Health Centre (primary care facility) and the larger, urban Mariakani Sub-county Hospital (secondary care facility). Both facilities would later participate in the PRECISE (Pregnancy Care Integrating Translational Science, Everywhere) pregnancy cohort study [20]. At the time of data collection, enrollment to the PRECISE cohort had not yet started.

ANC, routine delivery care, and emergency care for pregnancy complications are provided primarily by nurses/nurse-midwives with support by clinical officers (nonphysician clinicians) at both facilities. At Mariakani Hospital, doctors and an obstetrician/gynecologist provide specialist services for high-risk pregnancies. Ultrasound services are not available routinely, but can be undertaken at Mariakani Hospital (and other private facilities) upon referral for pregnancy complications or uncertainties regarding GA.

Study Participants and Sampling Methods

We sought to enroll two main groups of participants: (1) HCWs directly involved in the provision of services for pregnant women, as well as managers and health administrators; and (2) community members, represented by pregnant women participating in ANC and their family members (partners, as well as the pregnant woman's parents and parents-in-law).

HCWs were purposely sampled to cover providers at the ANC clinic, maternity, outpatient, and radiology (including ultrasound) departments. Health administrators were also purposely sampled to ensure inclusion of facility and subcounty managers overseeing reproductive health services. Pregnant women were approached by research assistants when they presented for routine ANC, and participating women could invite their partners or parents.

Data Collection

Data were collected between March and May 2019 by two Kenyan researchers: a social scientist (PMM) and a maternal health researcher and obstetrician (AK), who were assisted by two trained local research assistants who took notes during the sessions. Researchers were familiar with the local setting and the Kenyan health care system, and were fluent in both English and Swahili. The research assistants were also fluent in Mijikenda. None of the data collectors were involved in the participants' clinical care; however, some HCWs had previous interactions with the two researchers as part of PRECISE study preparations.

In-depth interviews (IDIs) with HCWs and focus group discussions (FGDs) with pregnant women and their families were conducted in person in private areas of the health facility, away from the clinical areas. We developed a semistructured interview guide, which was piloted on two HCWs at Rabai Health Centre and revised prior to the subsequent interviews. The topic guide began with simple assessment of prior exposure to computers, smartphones, and obstetric ultrasound, followed by a discussion on existing methods of assessing GA and the potential value to pregnant women and HCWs.

The study was started before clinical implementation of TraCer. A video demonstrating its use during routine ANC was shown to participants. The 5-minute video, recorded at one of the facilities in Swahili, showed a nurse using TraCer with a pregnant woman who was unsure of her LMP. In the video, the nurse shows the mother the image of the fetal heartbeat and the head of the baby on the tablet screen, reports the GA, and then gives a date for the next clinic visit.

Participants were encouraged to voice their thoughts and ask any questions during and immediately after watching the video. Participants were asked what they liked or disliked about TraCer as seen in the video, whether TraCer could be introduced to their health facility, how confident they would be in its findings, any outcomes (positive or negative) they expected with its introduction, the type of provider they thought could use TraCer, and whether they would recommend it to other pregnant women and health facilities.

IDIs were conducted in the participants' language of choice. HCWs preferred English, whereas pregnant women and their families preferred Swahili. All sessions were audio-recorded with permission and field notes were taken during each session. After each IDI and FGD, the research team debriefed to update field notes, discussed revisions and additional probes to the topic guide, and assessed data saturation.

Data Analysis

All recordings were transcribed verbatim in the language of the interview and translated to English (where applicable) by research assistants. A sample of transcripts was compared with the recordings to ensure accuracy. NVivo 12 (QSR International, Melbourne, Australia) was used to manage the transcripts, and to facilitate coding and collaborative data analysis. The data analysis team comprised three Kenyan researchers familiar with the study site and local languages (AK, PMM, GMM), including two who had participated in the data collection (AK, PMM) and two experienced Canadian social scientists (MWK, MV). The data analysis team first familiarized themselves with the transcripts. Employing a directed content analysis approach [21], the coding framework was developed deductively from the research question based on pre-existing definitions of acceptability and feasibility [22] that were modified to fit our study. *Acceptability* was assessed according to the perceptions of the appropriateness of TraCer to participant needs, preferences, and sociocultural norms, along with factors that would influence willingness to use the device. *Feasibility* was assessed according to perceptions on whether TraCer could be implemented in the study health facilities and factors required for its successful implementation.

Major themes and subthemes were explored related to acceptability and feasibility. The coding framework (see Multimedia Appendix 1) was tested on three transcripts to refine and ensure agreement between coders. Transcripts were then divided between coders for analysis. Emergent common and divergent patterns of responses between participants were explored through discussion within the team. Factors that were considered included differences in the site characteristics (urban vs rural, level of facility, access to ultrasound), HCW characteristics (skill level/cadre and prior experience of ultrasound), and community member characteristics (age, gender).

Ethical Considerations

The study obtained ethical approval from Aga Khan University Institutional Ethical Research Committee (2018_REC_47), King's College London (Ref HR-17/18-7855), and University of British Columbia (H18-02828). All participants provided individual written informed consent prior to research activities. Confidentiality and safe storage of the data were ensured through deidentification of transcripts and electronic storage in password-protected devices accessible only to members of the research team.

Results

Characteristics of the Study Sample

In total, we conducted 18 IDIs and 6 FGDs involving 52 community members; the IDIs lasted 27-64 minutes and the FGDs lasted 34-77 minutes. In the IDIs, nurses represented the largest group interviewed, and other cadres were clinical officers, doctors, sonographers, and public health officials (Table 1). Seven of these HCWs also had administrative responsibilities. The 52 members engaged in the FGDs included 31 in three FGDs at Rabai Health Centre and 21 in three FGDs at Mariakani Hospital. Overall, FGDs engaged 19 pregnant women, 15 partners (all male), and 18 parents (all mothers or mothers-in-law). Fewer than 20% (9/52) of pregnant women or their families had prior experience of ultrasound (Table 2).

Table 1. Characteristics of health care workers participating in interviews (N=18).

Characteristics	Health care workers, n
Site	
Rabai (rural)	8
Mariakani (urban)	10
Gender	
Male	7
Female	11
Profession	
Nurses	8
Clinical officers	4
Doctors	3
Others	3
Administrative responsibilities	
Yes	7
No	11
Age group (years)	
<35	7
35-44	6
45+	5
Prior exposure to ultrasound (observing)	
Yes	14
No	4
Prior exposure to ultrasound (performing)	
Yes	3
No	15

Table 2. Characteristics of pregnant women and their families participating in focus group discussions (N=52).

Characteristics	Pregnant women and their families, n
Site	
Rabai (rural)	31
Mariakani (urban)	21
Gender	
Men	15
Women	37
Category	
Pregnant women	19
Partners	15
Parents and parents-in-law	18
Age group (years)	
<35	30
35-44	8
45+	14
Prior exposure to ultrasound[a]	
Yes	9
No	43

[a]Refers to any exposure to ultrasound, including observing a procedure or having the procedure performed on them.

Acceptability

Acceptance

Participants' initial reactions were overwhelmingly positive. HCWs, pregnant women, and their family members stated that introduction of the tool to health facilities should be done as soon as possible, that they would recommend it to others, and that the introduction would encourage more women to come to the clinic earlier in their pregnancies. The expected high acceptance of TraCer in the community was predicted to result in increased ANC uptake and attendance throughout pregnancy.

> *I think it's good if you introduce it. It will help us to get the exact dates. Especially the mothers we are dealing with in the community, some of them will tell you that… "I can't tell when my last periods were. I just realized that am pregnant.*" [ANC clinic nurse IDI, Mariakani]
>
> *First, if the tool is brought, it will give the women motivation to come to the clinic. They will be desiring to come to the clinic…*
>
> *They will feel happy…because that tool is there. "You mean if I go to the clinic, I can see how my baby is doing?"*
>
> *It [TraCer] will make you come. You won't be saying let me wait for 2 months…*
>
> *You will come early* [Discussion in the pregnant women FGD, Rabai]

Pregnant women felt that seeing the baby would give them an indication of the baby's well-being and reduce their anxiety. The value of obtaining the GA would help them know the estimated date of delivery with more certainty. Other favorable features were the device's safety to both mother and baby, and the short duration of the assessment. In particular, participants from rural areas suggested that TraCer would reduce the need to travel to the urban referral site to access ultrasound services, resulting in savings in both time and money.

Pregnant women and their families expressed that they would trust the results provided by TraCer. The ability for one to see the image "for themselves" during use of TraCer was emphasized as a major contributor to trust. This was contrasted to other clinical procedures such as listening to the fetal heartbeat using a fetoscope, which is assessed only by the health provider. A printout of the image after the procedure was suggested. Trust in the findings of TraCer also reflected pre-existing trust in HCWs in general. Participants assumed the device would already have been tested prior to introduction to ensure its efficacy and safety.

> *…Detecting the baby's heartbeat…It means that the baby is alive…It means that he is doing well. So…it brings that confidence to the mother.* [Clinical Officer IDI, Rabai]
>
> *In my view, it will have reduced costs. Instead of traveling from here to Mariakani to queue there for a scan, and maybe I don't have the ability to go there because of my pocket [ability to pay].* [Male partners FGD, Rabai]
>
> *Because I will be watching from the tablet when am being examined…because you are seeing, definitely*

you will trust the results. [Mothers and Mothers-in-law FGD, Mariakani]

HCWs liked that TraCer would make it easier to obtain the GA, particularly for women with uncertain or unknown LMP. The ability to use new technology and give accurate GA estimation would boost their professional self-image and confidence in their services. The automated estimation of GA was viewed favorably, as it was perceived to be less susceptible to human error or interference by users. By providing simple GA assessment for all mothers in ANC clinics, TraCer would reduce workload at the few ultrasound facilities available because women requiring GA assessment only would not need a referral.

Overall, HCWs were optimistic that TraCer would find acceptance among pregnant women and community members. It would increase patient confidence in health providers and in their services, and would save them time and money spent traveling to referral facilities for ultrasound, ultimately leading to better retention of patients and continuity of care. HCWs expressed that many patients preferred health facilities that used advanced technologies, equating this to quality care.

…when we just want know the gestation age and fetal vitals it becomes more convenient than the ultrasound. [Doctor IDI, Mariakani]

Most people like to work in a place where you feel good. You feel good working because there are machines, there are less challenges. …At least when handling a machine, like for me who I've not used an ultrasound… Aaah, I feel motivated. I feel I've arrived… [Clinical Officer IDI, Rabai]

…we only have one ultrasound machine and only that one covering the whole hospital. Maternity cases, the wards, emergency department, outpatient clinics…and all patients to line up. And we only have only one sonographer doing all that. So it's kind of overwhelming to the sonographer when you line up maternity patients only to know the gestational age. [Other HCW IDI, Mariakani]

Hesitancy and Refusal

Although perceptions of TraCer were largely positive, some participants expressed preconditions that would need to be met before acceptance. The most frequent was the need to test the device's accuracy, often through comparison with other GA assessment methods. Among HCWs, this meant a comparison with formal ultrasound, while community members suggested comparing the estimated date of delivery obtained from TraCer with the actual delivery date. Concerns about the performance and accuracy of TraCer were raised more often by urban than rural participants. Among HCWs, these concerns were more common among the higher skilled providers (ie, doctors and sonographers). HCWs also expressed that the procedure would have to be brief to avoid delays, which would result in rejection of its use.

Despite the general perception that health technology was viewed favorably, some participants felt such a new innovation would not easily be trusted, and that observing how TraCer performs over time before fully trusting results was necessary. Others suggested that fear and uncertainty regarding safety for the fetus could result in opting against use. Consequently, it was suggested that a principle of informed choice be practiced to ensure pregnant women were able to choose whether or not to be examined using the device.

Cause it's something new, yeah. You can't trust something new without actually trying it out. You have to put it to the test. [Doctor IDI, Mariakani]

I'll take the gestation age using the ultrasound machine then we just compare. We will compare. Because for now I just don't know how accurate it [TraCer] is. [Other HCW IDI, Mariakani]

…I will compare because, you know, the TraCer machine you have said it measures the Cerebellum. … And in our case, in the ultrasound we don't use that. ...I have never used it. … I don't know how accurate it is. But, maybe with time. [Other HCW IDI, Mariakani]

You know, we Kenyans, when something new is introduced, we wait till the product has worked for a while. That's when we start appreciating it [Male partners FGD, Rabai]

Feasibility

Proposed Approach to Clinical Implementation

HCWs, pregnant women, and their families described ANC clinics as the ideal location for clinical implementation, as it is the typical first point of contact with most pregnant women at the health facility. This would also contribute to consolidation of services for pregnant women at a single location, reducing movement of pregnant mothers within the facility and saving time. Some suggested that TraCer be placed permanently at the ANC clinic to ensure it is always available to the women.

Participants unanimously agreed that TraCer should be provided by the ANC nurse as the nurse is already the main provider of ANC and is believed to have the relevant background knowledge of reproductive anatomy. In addition, some doctors and clinical officers felt that they too should be familiar with the tool since they also assessed pregnant women with complications in outpatient and maternity units.

…when it is introduced, it has to be based at the ANC [antenatal care clinic] because all our new mothers coming for antenatal, that is the first place they come. [Other HCW IDI, Mariakani]

I think all people who will be interacting with the pregnant mothers should be in position to use the tool effectively 'cause we interact with the pregnant mothers at different levels and I think when we are able to attend to them, any place, anytime, that would be better off than just to be basing it at the maternal child health care clinic only. [Clinical Officer IDI, Mariakani]

The need for training was emphasized by both community members and HCWs. All HCWs, regardless of prior training or education, felt confident that they would be able to use TraCer after appropriate training. There were contrasting views on the

nature and intensity of the training needed. Most HCWs felt that because they had basic knowledge of anatomy and physiology, on-job training of a short duration focusing on how to use and how to interpret the results would suffice. However, two doctors suggested that more intensive training was necessary to address potential difficulties in locating the baby's head in unusual fetal positions or in cases of multiple pregnancy. An additional recommendation was that other HCWs at the facility should be informed about the tool and its purpose to allow accurate communication to patients and community members.

> *…the nurse already [has] detailed [knowledge of the] anatomy of the woman's abdomen and also when it comes to the uterus. So I think with that background information they have in their profession practice, I think it won't be too difficult for it to require a very intensive training...* [Other HCW IDI, Rabai]

Concerns Regarding Implementation

Sustainable and consistent implementation were considered to be important by HCWs. HCWs raised concerns about a reliable supply of consumables needed and equipment maintenance. The device's hardware was often viewed favorably as "portable," "cordless," "lightweight," and "easy to carry around," in contrast to conventional ultrasound machines. However, some HCWs highlighted the risks of theft or misplacing the device because of its small size and portability. If TraCer became unavailable when mothers came to expect it, they would be less inclined to return to the clinic. To achieve service consistency, there would need to be adequate staffing and clear policies and guidelines for its use.

> *Then those supplies, who is maintaining those supplies?…who is going to sustain? So who continues to supply this?…* [Maternity nurse IDI, Mariakani]
>
> *…the facility that is using the device, can they be able to maintain the device? In case of any breakdown…and are we going to have some trained personnel service the device in case it becomes faulty?…that's what I was referring to as sustainability.* [ANC clinic nurse IDI, Rabai]

Misconceptions about TraCer were prevalent, especially among pregnant women and their families, but were also mentioned in 8 out of 18 interviews with HCWs. Many cited expectations of the device that were beyond its scope: the most frequent was that TraCer would provide detailed information about the well-being of the fetus and the mother. Other misconceptions included that the device would reveal fetal sex, confirm paternity, predict the exact date of delivery, that it could be used in the management of complications during labor and delivery, or that it could replace a formal obstetric ultrasound.

> *With that device, if a woman is pregnant, you will know everything that is happening insider her womb.* [Male partners FGD, Mariakani]
>
> *But with this one, I think there will be no need for the ultrasound because everything we'll see it at the antenatal clinic.* [ANC clinic nurse IDI, Rabai]

Some participants understood the role of TraCer as primarily to provide GA estimation but expressed their desire for the tool to give more information.

> *…it will be good also if more research is done so that we know the state of the mother and the baby, how old her pregnancy is and how to take care of herself so that she has a safe delivery. This tool also to be guiding this mother on how to take care of herself it will be much better. It will be good.* [Male partners FGD, Mariakani]

Discussion

Summary of Findings

In this study, we assessed a low-cost, portable, and artificial intelligence (AI)-enabled ultrasound device in two Kenyan settings. Overall, the device concept was found to be highly acceptable to women, their families, and HCWs, and implementation at health care facilities was felt to be feasible. The introduction of the device was predicted to contribute to reduced anxiety around the stage of pregnancy and fetal well-being, increase ANC clinic attendance, increase confidence by women in HCWs, and potential time- and cost-savings. It was also predicted to boost HCWs' professional self-image. The preferred users of TraCer were ANC nurses. There was a clear message that effective implementation requires adequate consideration on training, sustainability, and consistency of service. In addition, addressing potential misconceptions is a clear outcome of this work.

Various features of TraCer contributed to its acceptance. In particular, the value of seeing an image of the baby and the ability to see fetal heart activity were appreciated and strongly linked to trust of the device's findings/results. The added value of "seeing" has been highlighted in other studies [23]. Device features such as the small size, portability, and absence of connecting cords were cited as strengths.

Previous studies conducted in sub-Saharan Africa highlight the positive attitude toward new technological innovations in health [24-26]. Despite little previous experience, HCWs showed high levels of acceptance toward new technologies owing to expectations that new tools would be easy to use, make work easier, and take less time [25,26]. This is also true for studies evaluating ultrasound devices [23,27,28]. Our study is unique in that it assessed a tool using AI to automate analysis, and that the tool is intended for a narrow use-case of GA assessment. Despite these features, the positive attitude toward the tool and the willingness to have it introduced was shared. The use of the specific device was viewed favorably and equated with quality, with HCWs expressing eagerness to learn its use. In our study, HCWs expressed confidence in automation and felt that the results would be more reliable due to reduction of human error and absence of interference from providers. This is an important finding relevant to future technologies incorporating AI in medical devices in these settings.

Some hesitancy, due to novelty, was also expressed, indicating the need for robust testing of performance and the effect of implementation over time. This is an important finding for

implementation, demonstrating that immediate acceptance of new tools upon introduction cannot be expected. It is likely that the experience during early phases of implementation may influence eventual acceptability of novel tools.

There is increasing use of low-cost portable ultrasound in low-resource settings [29]; however, the majority of studies have been conducted in large tertiary hospitals with only a few conducted in rural facilities. Our study demonstrated higher acceptance of TraCer in the rural site. It is possible that this is due to the lack of any ultrasound service in that setting (Rabai) and the challenges associated with referral to the urban site. This unequal distribution of ultrasound services is common in sub-Saharan Africa [13], and sites for future implementation of TraCer and similar technologies should be carefully selected with this in mind. In contrast, the higher hesitancy in the urban site (Mariakani) could be a reflection of the higher existing technical skills and better understanding of potential pitfalls. Thus, HCWs with higher skill levels and prior experience with pregnancy ultrasound had more questions regarding accuracy and recommended more intensive training. They should therefore be provided with robust technical information and given ample time for training and testing.

This study suggests that implementation of TraCer in this setting will be feasible, but several potential barriers to effective implementation should be considered. Previous studies have demonstrated the feasibility of the use of small portable (compact) ultrasound devices, particularly for obstetric applications [30]. Similar to other studies [15], HCWs highlighted the need for equipment maintenance to ensure sustainability and consistency of the service. Service interruptions may be disappointing to staff and pregnant women and may affect their attitudes toward the tool [25].

HCWs were confident that they would be able to perform the procedure if given adequate training, but their views on the desired nature and duration of the training were varied. Other studies have evaluated combinations of didactics courses, hands-on instruction, supervised scanning, and lectures [31,32]. The duration of training for obstetric point-of-care studies in one review [31] ranged from 3 days to 3 months. Findings from our study are not sufficient to inform the design of a training package for TraCer, which will be evaluated during implementation. Addition of a training module on basic troubleshooting and repair of the device may be beneficial.

Like in other settings, misconceptions around the use of the technological innovations was common [28], with frequent reports of overestimation of its diagnostic potential [28,33]. It is possible that part of the positive attitudes toward TraCer reported in this study may have been influenced by these misconceptions and overestimation of the tool's scope and function. Providing false reassurance of fetal and maternal well-being could contribute to delays in care-seeking for women, or may lead to blame in case of an adverse pregnancy outcome. Utmost care should be exercised to avoid HCWs becoming less thorough during clinical assessments or changing referral thresholds for further care due to such false reassurance of fetal well-being [27,33]. Thus, the role and scope of TraCer should be emphasized for both pregnant women and HCWs. Referral pathways for formal ultrasound assessment should remain open and indications for this made clear.

Nevertheless, the presence of a visible fetal heart beat is an important measure of fetal well-being. HCWs perceived that using TraCer would make assessment of the fetal heartbeat faster and easier to operate than a Pinard fetoscope or Doppler fetal monitor. We suggest that messaging around fetal status should acknowledge the confirmation that the baby is alive at the time of the examination, but clarify that this does not give reassurance of well-being.

Proposed Structure of Implementation of TraCer

In view of the above findings, we propose that a fully validated and AI-driven TraCer device be used by nurses at the ANC clinic for assessment of all pregnant women. In this setting, TraCer would assess fetal viability and provide an accurate estimate of GA. Where there are indications for full obstetric ultrasound or other concerns for fetal well-being, women should be referred and encouraged to proceed for formal ultrasound assessment in line with current practice. With this approach, more women would have an accurate GA to inform subsequent decision-making and scheduling of effective ANC. Automated GA assessment using TraCer could reduce referrals for GA assessment only, giving higher-skilled ultrasound providers more time to focus on women needing specialist assessments. Such an approach has been suggested by other studies [13,34,35].

Strengths and Limitations

Our study has a number of strengths. Data were collected from various participant groups and evaluated in the setting intended for use (ie, low-resource settings in LMICs). To increase the generalizability of findings, data were obtained from two health care facilities with differences in location (periurban vs rural), level (primary vs secondary), and availability of ultrasound services. The research team comprised skilled qualitative researchers and experts in maternal health and pregnancy ultrasound. The data collection and analysis team included researchers familiar with the local language and context.

The use of video role-play has been successful in eliciting rich data on acceptability and feasibility of the tool prior to clinical implementation. This approach has allowed an early evaluation, ensuring that the initial views of users are incorporated into further tool development and planning for actual clinical implementation. The study has also helped to understand baseline user expectations that can be used to measure clinical implementation indicators.

Although useful, video role-play also has limitations, as it may create an inaccurate perception of proposed clinical implementation. In this case, study participants felt that use of the device took very little time, an impression that could be created by an edited video. To overcome this potential limitation, further assessments during active clinical implementation of the tool are important.

Some of the research team had had interactions with study participants (HCWs) prior to data collection as part of preparations for a larger research study. This is important as it

could have introduced some social desirability bias, with these responders expressing a more favorable view toward TraCer. This was mitigated by selecting the interviewer who had had the least prior interaction with the participant.

Conclusion

We have shown that there is potential to implement AI-enabled ultrasound innovations in low-income settings, including a device that offers only a selected fetal assessment (in this case, GA). It is highly likely that the TraCer tool can be implemented in this and similar settings, and that users will find it valuable. Further device developments should ensure that the tool is simple and easy to use, and that results are obtained within a short time frame. Prior to clinical implementation, robust accuracy data must be provided and measures should be taken to ensure sustainability and consistency of the service. Clarity of messaging about the tool and its role in pregnancy is essential to prevent misconceptions and misuse. Our study points to further assessments required during the subsequent phases of clinical implementation, feasibility, acceptability, and device usability, as well as clinical validity and safety of the tool.

Acknowledgments

This project was supported by a grant to the PRECISE Network, funded by the UK Research and Innovation Grand Challenges Research Fund GROW Award scheme (grant number MR/P027938/1) and the Bill & Melinda Gates Foundation Grand Challenges Explorations Program (grant number OPP1128591). ATP is supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the National Institute of Health Research Biomedical Research Centre funding scheme. The funding sources had no involvement at any stage of research from design, conduct, to writing.

Authors' Contributions

The members of The PRECISE Network are as follows: Umberto D'Alessandro, Anna Roca, Hawanatu Jah, Andrew Prentice, Melisa Martinez-Alvarez, Brahima Diallo, Abdul Sesey, Baboucarr Njie, Fatima Touray, Yahaya Idris, Fatoumata Kongira, and Mathurin Diatta (The London School of Hygiene and Tropical Medicine, Fajara, Gambia); Marleen Temmerman, Angela Koech, Sikolia Wanyonyi, Patricia Okiro, Geoffrey Omuse, Grace Mwashigadi, Moses Mukhanya, Onesmus Wanje, Consolata Juma, Joseph Mutunga, Marvine Ochieng, Claire Ngure, Peris Musitia, Claire Ngure, Joy Njoroge, Nathan Barreh, Mwanajuma Bakari, Francis Kazungu, Lydia Mutinda, Mary Mwania, and Sophie Onsare (Aga Khan University, Nairobi, Kenya); Manhiça, Manhiça, Esperança Sevene, Corssino Tchavana, Salesio Macuacua, Anifa Vala, Helena Boene, Lazaro Quimice, Sonia Maculuve, Ana Ilda Biza, Ernesto Maximiano, and Silvestre Cutana (Centro de Investigação em Saúde de Manhiça, Manhiça, Mozambique); Peter von Dadelszen, Laura A. Magee, Rachel Craik, Hiten Mistry, Marie-Laure Volvert, Cristina Escalona, Anne Rerimoi, and Giulia Ghillia (Central Co-ordinating Team, Department of Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London); Donna Russell (Donna Russell Consulting); Prestige Tatenda Makanga, Liberty Makacha, Reason Mlambo, Shirley Chapunza, and Tendai Nkomo (Midlands State University, Zimbabwe); Lucilla Poston, Jane Sandall, Rachel Tribe, Andrew Shennan, Sophie Moore, and Tatiana Salisbury (Kings College London); Aris Papageorghiou, Alison Noble, Rachel Craik, Yuan Goa, Sridevi Beriwal, Lee Lok Hin, Elizabeth Bradburn, Alice Self, and Jayne Lander (University of Oxford); Hannah Blencowe, Veronique Filippi, Joy Lawn, Matt Silver, and Ursula Gazeley (London School of Hygiene and Tropical Medicine); Judith Cartwright, Guy Whitley, and Sanjeev Krishna (St George's, University of London); Marianne Vidler, Jing (Larry) Li, Jeffrey N Bone, Mai-Lei (Maggie) Woo Kinshella, Domena Tu, Ash Sandhu, and Kelly Pickerill (University of British Columbia); Chris Clarke (C-Squared Development); and Ben Barratt (Imperial College London).

Conflicts of Interest

JAN and ATP are Senior Advisors of Intelligent Ultrasound. All other authors declare no competing interests.

Multimedia Appendix 1
Codebook for data analysis.
[DOCX File , 17 KB - humanfactors_v9i4e34823_app1.docx]

References

1. Kramer MS. Born too small or too soon. Lancet Glob Health 2013 Jul;1(1):e7-e8 [FREE Full text] [doi: 10.1016/S2214-109X(13)70014-7] [Medline: 25103586]
2. Preterm labour and birth NICE guideline [NG25]. National Institute for Health Care and Excellence. 2015. URL: https://www.nice.org.uk/guidance/ng25 [accessed 2021-11-02]
3. Yuce T, Keskin M, Seval MM, Söylemez F. Effect of the timing of delivery on perinatal outcomes at gestational hypertension. Interv Med Appl Sci 2015 Jun;7(2):59-62 [FREE Full text] [doi: 10.1556/1646.7.2015.2.3] [Medline: 26120477]

4. Wylie BJ, Kalilani-Phiri L, Madanitsa M, Membe G, Nyirenda O, Mawindo P, et al. Gestational age assessment in malaria pregnancy cohorts: a prospective ultrasound demonstration project in Malawi. Malar J 2013 Jun 04;12(1):183 [FREE Full text] [doi: 10.1186/1475-2875-12-183] [Medline: 23734718]
5. Jackson K, Harrington JW. SGA and VLBW infants: outcomes and care. Pediatr Rev 2018 Jul;39(7):375-377. [doi: 10.1542/pir.2017-0091] [Medline: 29967087]
6. Kullinger M, Haglund B, Kieler H, Skalkidou A. Effects of ultrasound pregnancy dating on neonatal morbidity in late preterm and early term male infants: a register-based cohort study. BMC Pregnancy Childbirth 2016 Oct 31;16(1):335 [FREE Full text] [doi: 10.1186/s12884-016-1129-z] [Medline: 27799069]
7. Kemp B, Koech A, Mwangudzah H, Seale A, Kennedy S, Berkley J, et al. OC23.02: To what extent does ultrasound estimation of gestational age affect the rate of preterm births in a rural population in sub-Saharan Africa (sSA)? In: Ultrasound Obstet Gynecol. 2013 Oct 01 Presented at: 23rd World Congress on Ultrasound in Obstetrics and Gynecology (ISUOG); October 6-9, 2013; Sydney, Australia p. 46. [doi: 10.1002/uog.12712]
8. Dietz PM, England LJ, Callaghan WM, Pearl M, Wier ML, Kharrazi M. A comparison of LMP-based and ultrasound-based estimates of gestational age using linked California livebirth and prenatal screening records. Paediatr Perinat Epidemiol 2007 Sep;21(Suppl 2):62-71 [FREE Full text] [doi: 10.1111/j.1365-3016.2007.00862.x] [Medline: 17803619]
9. Antenatal care for uncomplicated pregnancies. Clinical guideline [CG62]. National Institute for Health Care and Excellence. 2008. URL: https://www.nice.org.uk/guidance/cg62 [accessed 2021-11-02]
10. Sarker BK, Rahman M, Rahman T, Rahman T, Rahman F, Khalil JJ, et al. Factors associated with calendar literacy and last menstrual period (LMP) recall: a prospective programmatic implication to maternal health in Bangladesh. BMJ Open 2020 Dec 13;10(12):e036994 [FREE Full text] [doi: 10.1136/bmjopen-2020-036994] [Medline: 33318107]
11. Government of Kenya. Kenya Demographic and Health Survey: Key Indicators 2014. reliefweb. 2015. URL: https://reliefweb.int/report/kenya/kenya-demographic-and-health-survey-key-indicators-2014 [accessed 2021-11-02]
12. de Onis M, Blössner M, Villar J. Levels and patterns of intrauterine growth retardation in developing countries. Eur J Clin Nutr 1998 Jan;52(Suppl 1):S5-S15 [FREE Full text] [Medline: 9511014]
13. Wanyonyi S, Mariara C, Vinayak S, Stones W. Opportunities and challenges in realizing universal access to obstetric ultrasound in sub-Saharan Africa. Ultrasound Int Open 2017 Apr 07;3(2):E52-E59 [FREE Full text] [doi: 10.1055/s-0043-103948] [Medline: 28596999]
14. Gerein N, Green A, Pearson S. The implications of shortages of health professionals for maternal health in sub-Saharan Africa. Reprod Health Matters 2006 May 19;14(27):40-50 [FREE Full text] [doi: 10.1016/s0968-8080(06)27225-2]
15. Shah S, Bellows BA, Adedipe AA, Totten JE, Backlund BH, Sajed D. Perceived barriers in the use of ultrasound in developing countries. Crit Ultrasound J 2015 Dec 19;7(1):28 [FREE Full text] [doi: 10.1186/s13089-015-0028-2] [Medline: 26123609]
16. Alliance for Maternal Newborn Health Improvement (AMANHI) Gestational Age Study Group. Simplified models to assess newborn gestational age in low-middle income countries: findings from a multicountry, prospective cohort study. BMJ Glob Health 2021 Sep 13;6(9):e005688 [FREE Full text] [doi: 10.1136/bmjgh-2021-005688] [Medline: 34518201]
17. Rodriguez-Sibaja MJ, Villar J, Ohuma EO, Napolitano R, Heyl S, Carvalho M, et al. Fetal cerebellar growth and Sylvian fissure maturation: international standards from Fetal Growth Longitudinal Study of INTERGROWTH-21 Project. Ultrasound Obstet Gynecol 2021 Apr;57(4):614-623 [FREE Full text] [doi: 10.1002/uog.22017] [Medline: 32196791]
18. Brooks A, Smith TA, de Savigny D, Lengeler C. Implementing new health interventions in developing countries: why do we lose a decade or more? BMC Public Health 2012 Aug 21;12(1):683 [FREE Full text] [doi: 10.1186/1471-2458-12-683] [Medline: 22908877]
19. Lunze K, Higgins-Steele A, Simen-Kapeu A, Vesel L, Kim J, Dickson K. Innovative approaches for improving maternal and newborn health--A landscape analysis. BMC Pregnancy Childbirth 2015 Dec 17;15(1):337 [FREE Full text] [doi: 10.1186/s12884-015-0784-9] [Medline: 26679709]
20. von Dadelszen P, Flint-O'Kane M, Poston L, Craik R, Russell D, Tribe RM, PRECISE Network. The PRECISE (PREgnancy Care Integrating translational Science, Everywhere) Network's first protocol: deep phenotyping in three sub-Saharan African countries. Reprod Health 2020 Apr 30;17(Suppl 1):51 [FREE Full text] [doi: 10.1186/s12978-020-0872-9] [Medline: 32354357]
21. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288 [FREE Full text] [doi: 10.1177/1049732305276687] [Medline: 16204405]
22. Peters D, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. BMJ 2013 Nov 20;347:f6753 [FREE Full text] [doi: 10.1136/bmj.f6753] [Medline: 24259324]
23. Oluoch DA, Mwangome N, Kemp B, Seale AC, Koech A, Papageorghiou AT, et al. "You cannot know if it's a baby or not a baby": uptake, provision and perceptions of antenatal care and routine antenatal ultrasound scanning in rural Kenya. BMC Pregnancy Childbirth 2015 May 29;15(1):127 [FREE Full text] [doi: 10.1186/s12884-015-0565-5] [Medline: 26021564]
24. Riaz A, Cambaco O, Ellington LE, Lenahan JL, Munguambe K, Mehmood U, et al. Feasibility, usability and acceptability of paediatric lung ultrasound among healthcare providers and caregivers for the diagnosis of childhood pneumonia in resource-constrained settings: a qualitative study. BMJ Open 2021 Mar 11;11(3):e042547 [FREE Full text] [doi: 10.1136/bmjopen-2020-042547] [Medline: 33707268]

25. Rothstein JD, Jennings L, Moorthy A, Yang F, Gee L, Romano K, et al. Qualitative assessment of the feasibility, usability, and acceptability of a mobile client data app for community-based maternal, neonatal, and child care in rural Ghana. Int J Telemed Appl 2016;2016:2515420. [doi: 10.1155/2016/2515420] [Medline: 28070186]
26. Sukums F, Mensah N, Mpembeni R, Kaltschmidt J, Haefeli WE, Blank A. Health workers' knowledge of and attitudes towards computer applications in rural African health facilities. Glob Health Action 2014 Oct 27;7(1):24534 [FREE Full text] [doi: 10.3402/gha.v7.24534] [Medline: 25361721]
27. Kim ET, Singh K, Moran A, Armbruster D, Kozuki N. Obstetric ultrasound use in low and middle income countries: a narrative review. Reprod Health 2018 Jul 20;15(1):129 [FREE Full text] [doi: 10.1186/s12978-018-0571-y] [Medline: 30029609]
28. Firth ER, Mlay P, Walker R, Sill PR. Pregnant women's beliefs, expectations and experiences of antenatal ultrasound in Northern Tanzania. Afr J Reprod Health 2011 Jun;15(2):91-107. [Medline: 22590896]
29. Stewart KA, Navarro SM, Kambala S, Tan G, Poondla R, Lederman S, et al. Trends in ultrasound use in low and middle income countries: a systematic review. Int J MCH AIDS 2020 Jan 03;9(1):103-120 [FREE Full text] [doi: 10.21106/ijma.294] [Medline: 32123634]
30. Harris R, Marks WM. Compact ultrasound for improving maternal and perinatal care in low-resource settings: review of the potential benefits, implementation challenges, and public health issues. J Ultrasound Med 2009 Aug;28(8):1067-1076. [doi: 10.7863/jum.2009.28.8.1067] [Medline: 19643790]
31. Abrokwa SK, Ruby LC, Heuvelings CC, Bélard S. Task shifting for point of care ultrasound in primary healthcare in low- and middle-income countries-a systematic review. EClinicalMedicine 2022 Mar;45:101333 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101333] [Medline: 35284806]
32. Viner AC, Okolo ID, Norman JE, Stock SJ, Reynolds RM. Training in ultrasound to determine gestational age in low- and middle-income countries: a systematic review. Front Glob Womens Health 2022;3:854198 [FREE Full text] [doi: 10.3389/fgwh.2022.854198] [Medline: 35368997]
33. Tautz S, Jahn A, Molokomme I, Görgen R. Between fear and relief: how rural pregnant women experience foetal ultrasound in a Botswana district hospital. Soc Sci Med 2000 Mar;50(5):689-701 [FREE Full text] [doi: 10.1016/s0277-9536(99)00321-4]
34. Swanson JO, Kawooya MG, Swanson DL, Hippe DS, Dungu-Matovu P, Nathan R. The diagnostic impact of limited, screening obstetric ultrasound when performed by midwives in rural Uganda. J Perinatol 2014 Jul 03;34(7):508-512 [FREE Full text] [doi: 10.1038/jp.2014.54] [Medline: 24699218]
35. Holmlund S, Ntaganira J, Edvardsson K, Lan PT, Semasaka Sengoma JP, Lesio Kidanto H, et al. Health professionals' experiences and views on obstetric ultrasound in Rwanda: a cross-sectional study. PLoS One 2018;13(12):e0208387 [FREE Full text] [doi: 10.1371/journal.pone.0208387] [Medline: 30513102]

Abbreviations

AI: artificial intelligence
ANC: antenatal care
FGD: focus group discussion
GA: gestational age
HC: head circumference
HCW: health care worker
IDI: in-depth interview
LMIC: low- and middle-income country
LMP: last menstrual period
PRECISE: Pregnancy Care Integrating Translational Science, Everywhere
TCD: transcerebellar diameter

Edited by A Kushniruk; submitted 01.12.21; peer-reviewed by C Taguibao, E Borycki; comments to author 02.08.22; revised version received 27.09.22; accepted 09.11.22; published 27.12.22.

Please cite as:
Koech A, Musitia PM, Mwashigadi GM, Kinshella MLW, Vidler M, Temmerman M, Craik R, von Dadelszen P, Noble JA, Papageorghiou AT, The PRECISE Network
Acceptability and Feasibility of a Low-Cost Device for Gestational Age Assessment in a Low-Resource Setting: Qualitative Study
JMIR Hum Factors 2022;9(4):e34823
URL: https://humanfactors.jmir.org/2022/4/e34823
doi:10.2196/34823
PMID:36574278

Original Paper

Understanding a Care Management System's Role in Influencing a Transitional-Aged Youth Program's Practice: Mixed Methods Study

Rubina F Rizvi[1], MD, PhD; Courtney B VanHouten[1], MA; Van C Willis[1], PhD; Bedda L Rosario[1], MPH, PhD; Brett R South[1], MS, PhD; Megan Sands-Lincoln[1], MPH, PhD; David Brotman[1], MS; Jeffery Lenert[1], MBA, MD; Jane L Snowdon[1], PhD; Gretchen P Jackson[1,2,3,4,5], MD, PhD

[1]IBM Watson Health, Cambridge, MA, United States

[2]Department of Pediatric Surgery, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States

[4]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[5]Intuitive Surgical, Sunnyvale, CA, United States

Corresponding Author:
David Brotman, MS
IBM Watson Health
75 Binney St
Cambridge, MA, 02142
United States
Phone: 1 4432865723
Email: david.brotman@ibm.com

Abstract

Background: Extended foster care programs help prepare transitional-aged youth (TAY) to step into adulthood and live independent lives. Aspiranet, one of California's largest social service organizations, used a social care management solution (SCMS) to meet TAY's needs.

Objective: We aimed to investigate the impact of an SCMS, IBM Watson Care Manager (WCM), in transforming foster program service delivery and improving TAY outcomes.

Methods: We used a mixed methods study design by collecting primary data from stakeholders through semistructured interviews in 2021 and by pulling secondary data from annual reports, system use logs, and data repositories from 2014 to 2021. Thematic analysis based on grounded theory was used to analyze qualitative data using NVivo software. Descriptive analysis of aggregated outcome metrics in the quantitative data was performed and compared across 2 periods: pre-SCMS implementation (before October 31, 2016) and post-SCMS implementation (November 1, 2016, and March 31, 2021).

Results: In total, 6 Aspiranet employees (4 leaders and 2 life coaches) were interviewed, with a median time of 56 (IQR 53-67) minutes. The majority (5/6, 83%) were female, over 30 years of age (median 37, IQR 32-39) with a median of 6 (IQR 5-10) years of experience at Aspiranet and overall field experience of 10 (IQR 7-14) years. Most (4/6, 67%) participants rated their technological skills as expert. Thematic analysis of participants' interview transcripts yielded 24 subthemes that were grouped into 6 superordinate themes: study context, the impact of the new tool, key strengths, commonly used features, expectations with WCM, and limitations and recommendations. The tool met users' initial expectations of streamlining tasks and adopting essential functionalities. Median satisfaction scores around pre- and post-WCM workflow processes remained constant between 2 life coaches (3.25, IQR 2.5-4); however, among leaders, post-WCM scores (median 4, IQR 4-5) were higher than pre-WCM scores (median 3, IQR 3-3). Across the 2 study phases, Aspiranet served 1641 TAY having consistent population demographics (median age of 18, IQR 18-19 years; female: 903/1641, 55.03%; race and ethnicity: Hispanic or Latino: 621/1641, 37.84%; Black: 470/1641, 28.64%; White: 397/1641, 24.19%; Other: 153/1641, 9.32%). Between the pre- and post-WCM period, there was an increase in full-time school enrollment (359/531, 67.6% to 833/1110, 75.04%) and a reduction in part-time school enrollment (61/531, 11.5% to 91/1110, 8.2%). The median number of days spent in the foster care program remained the same (247, IQR 125-468 years); however, the number of incidents reported monthly per hundred youth showed a steady decline, even with an exponentially increasing number of enrolled youth and incidents.

Conclusions: The SCMS for coordinating care and delivering tailored services to TAY streamlined Aspiranet's workflows and processes and positively impacted youth outcomes. Further enhancements are needed to better align with user and youth needs.

(JMIR Hum Factors 2022;9(4):e39646) doi:10.2196/39646

KEYWORDS

care management solution; foster care youth; mixed methods study; interviews; qualitative data; quantitative data; process improvement

Introduction

Foster Care

Extended foster care is considered an effective intervention to help prepare young adults to transition into adulthood to live independent, healthy, and successful lives and to help mitigate the risk of undesirable outcomes. However, providing extended foster care is a complex process and requires coordination and collaboration among various social service departments and stakeholders. The *complex social reality* of problems encountered in social work is seen universally [1]. Often, youths in foster care face multiple issues at a single point in time with competing priorities around when and how to address them [2]. A robust and integrated system of care is needed to understand the optimal processes and services to provide support for each individual youth with unique needs [3]. According to child welfare policy and services experts, the complex task of managing youth in foster care has been considered similar to the role of a parent and hence called "corporate parenting" [4].

The transition to adulthood is a critical stage in one's life. A young adult may encounter multiple significant challenges in accessing the required resources and developing the skills essential to becoming self-sufficient. Transitional-aged youths (TAYs) are young people between the ages of 16 and 24 years [5] who are in transition from state custody or foster care environments and are at risk. Such individuals are at high risk of experiencing several undesirable outcomes, such as mental health or behavioral disorders [6], economic insecurity [7], housing instabilities [8,9], criminal justice involvement [8], and unintended pregnancies [10]. Aspiranet, one of California's largest social service organizations with headquarters in San Francisco, California, provides a bridge between foster care and independence by providing extended foster care services to enable TAY with the necessary life skills, counseling, and medical treatment. Extended foster care can help protect TAY from some of these adverse outcomes [11].

Foster Care Challenges

Foster program–specific challenges, such as large program sizes, high employee and benefit recipient turnover, disjointed care coordination, and complex health and emotional needs in foster children, make this already challenging task especially difficult. In the United States ≥400,000 children and youths are in the foster care system, with more than 150,000 youths aged 14 to 21 years enrolled in 2017 [12]. According to the Adoption and Foster Care Analysis and Reporting System data, approximately 20,000 youths age out of foster programs each year [13]. In California, the foster care population is the largest in any state, with approximately 51,000 children in foster care and a foster care rate of 5.8 per 1000 children using the 2019 US Census Bureau data, which further compounds the challenge [14]. TAYs in foster care programs need tailored attention and action plans to address their individual and often multiple needs according to the California Youth Transitions to Adulthood Study [15]. Overall, almost one-third (30.9%) of the youth reported being homeless sometime during the 5-year study period [16]. Almost 53% (n=404) of California Youth Transitions to Adulthood Study participants were found to have a positive diagnosis for one or more current mental and behavioral health disorder, with a greater frequency of diagnosis among females than males (57.5% and 46.9%, respectively) [17]. Overall, about 44% of youths had justice system involvement before the age 17 interview [18].

The complexity of social care services among extended foster care youth, along with siloed services and a lack of integrated information systems, leads to less-effective and less-efficient care delivery and communication [19]. To address this gap, the establishment of data governance and management, streamlining of processes, and integration of disparate systems under a common platform are needed [20,21].

Care Management Solutions

Care management programs and systems have been widely adopted in health care provider organizations, and emerging evidence suggests that technologies to support care management may provide value in delivering complex care across social services outside the clinical environment. As reported in earlier studies, an advanced technology-based tool could enhance the care delivery process through the integration of data- and information-coordinated care. The benefit of such a tool ranges from providing comprehensive services to various stakeholders across governments, nonprofits, and other organizations to addressing the challenges of information silos and limited resources [19,22].

Although evidence supports the value of case management systems in stabilizing vulnerable adult citizens, to our knowledge, such technologies have not yet been examined in the TAY population. We sought to investigate whether and how a cloud-based social care management solution could improve the care of transitional youth in a large foster care program using a mixed methods approach. We hypothesized that the care management system would not only streamline workflows and facilitate social care service providers in delivering more effective and efficient care but also improve youth outcomes and successful transition into a self-sufficient adulthood.

Methods

Study Site

Aspiranet is one of California's largest nonprofit social service organizations providing support for children, youth, and families [23,24]. Founded in 1975 with a single group home, Aspiranet now runs 44 community-based sites across the state. It provides a spectrum of services such as foster care and adoption support, residential group home care, support for youth making the transition from foster care to adulthood, mental and behavioral health services, intensive home-based care, and community-based family resources [25]. More than 35,000 children, youths, and parents have benefited from Aspiranet's services.

Care Management System

In 2016, Aspiranet adopted IBM Watson Care Manager (WCM) to transform its social services programs from a traditional, paper-based process to an integrated and interdisciplinary technology-based system. WCM and IBM Health and Human Services Connect360 (data hub) is an integrated and configurable cloud-based software-as-a-service solution that allows teams to collaborate across agencies and reduce silos and provides a holistic view of the person from a historical and program perspective [26].

WCM enables child welfare workers to develop integrated care plans, triage vulnerable residents to the services, and foster collaboration with other care professionals (eg, within and external to Aspiranet, such as outpatient and school-based mental health services) to optimize service delivery. WCM integrates the Casey Life Skills (CLS) tool kit to help guide and assess the independent skills that youths need to achieve their long-term goals [27]. CLS assesses life skill domains built on the principles of psychometric measurement that defines a life skills assessment not as behavioral performance but as a general life skills ability. Aspiranet fully implemented WCM over the course of 2 years (2016-2018).

Study Design and End Points

Overview

A mixed methods study was conducted to understand the impact of WCM on Aspiranet's workflow processes and youth outcomes. The primary study end points were both qualitative and quantitative measures. Qualitative examples included Aspiranet employees' perspectives on the impact of new technology on care management effectiveness and efficiencies compared with previous practices. We also looked at quantitative measures including program outcomes such as a change in school enrollment, employment and housing status, number of undesired events, and user satisfaction scores (5-point Likert scale) pre- and postimplementation. Qualitative data were obtained from interviews with Aspiranet employees who were WCM users. Quantitative data were extracted from reports, backend use logs, and data repositories. The study was conducted from January 1, 2014, to March 31, 2021, with data collection divided into 2 main phases: before using WCM (before October 31, 2016) and after using WCM (November 1, 2016, to March 31, 2021).

Study Participants

Purposive sampling in collaboration with Aspiranet administrative staff was used to recruit Aspiranet employees (ie, leaders and life coaches) to participate in the study. Aspiranet leaders were executives in various roles, and life coaches were the case managers who provided supportive services to the at-risk youth transitioning to adulthood at various Aspiranet locations. All participants were aged ≥18 years, who had experience in administering the TAY program both before and after the implementation of WCM. Informed consent was obtained from each participant.

Ethics Approval

All study participants provided written consent to participate in personal audio-recorded interviews. This study was deemed exempt from human subjects research regulations by the Western Institutional Review Board (WCG IRB Work Order #1-1437184-1), exempted under 45 CFR § 46.104(d). All key study personnel involved in the design or conduct of the research completed the required education and certification on the protection of human research subjects.

Qualitative Data: Semistructured Interviews

In-depth interviews were conducted over videoconferencing on WebEx using a semistructured topic guide (Multimedia Appendix 1) to elicit information about fostering collaboration, optimizing care management for TAY, workflows, the experience with WCM and legacy systems, benefits and challenges of WCM, and the impact on efficiency and youth-centered care coordination. Interviews were conducted by a physician informatician (RFR) and medical ethnographer (CBV) with training and expertise in evaluating user experience and the effects of health information technologies.

Quantitative Data: Program Outcomes

Overview

Pre-WCM secondary data were primarily collected from the Aspiranet Legacy Data Hub containing higher-lever descriptive data about employees and enrolled youth. Owing to inherent barriers associated with management and tracking, no paper-based data were collected. Post-WCM data were collected from both the Aspiranet Legacy Data Hub and a Cognos reporting system (post-WCM implementation), describing a more granular picture of WCM use and program outcomes; for example, goals, actions, and tasks. For the data coming from legacy databases, the technical staff at Aspiranet performed data extraction, cleansing, and quality checks, which were validated by 2 IBM researchers (BRS and RFR). For the Cognos data, an IBM IT specialist worked closely with the Aspiranet staff to run the reports. Deidentified backend data for 2 specific periods, pre- and post-WCM implementation, were obtained for analysis.

The data collected included the number of enrolled TAY, the number of goals established (ie, a set achievement to reach a milestone, eg, getting a driver's license or enrolling in college), tasks (ie, actions taken to reach that specific goal, eg, pass a permit test or file a college application), and the number of incidents or undesired events (eg, car accident, school suspension, and police involvement).

Qualitative Analysis

Interview transcripts were transcribed, deidentified, and analyzed by 2 members of the research team (RFR and CBV) using NVivo software (QSR International) [28,29]. A thematic approach directed by grounded theory [29] was used to develop a codebook in which disagreements were resolved through discussion until consensus was reached.

Coding in qualitative analysis is a process of systematically categorizing qualitative data into themes and patterns. The process entails open coding (ie, first breaking down textual data into discrete parts), followed by axial coding (ie, drawing connections between codes), and finally selective coding (ie, cataloging all categories together around 1 core theme) [30]. Coding was performed sequentially by 2 research team members (RFR and CBV), with regular communication and discussion to ensure consistency and accuracy. The perspectives of leaders and life coaches were collected about pre- and post-WCM experiences, such as the impact on workflows, quality of service delivery, and youth outcomes.

Quantitative Analysis

Descriptive statistics for each outcome of interest were summarized as frequencies (percentages) for categorical data or as mean (SD) or median (IQR) for normally distributed or nonnormally distributed continuous data, as appropriate for pre- and post-WCM periods. Examination of normal distribution assumptions for continuous data was determined using quantile-quantile plots and histograms. These plots are graphical ways to check whether the data follow a normal distribution. All statistical analyses were conducted using SAS language on WPS Analytics (version 4.2; World Programming) [31].

Results

Overview

The mixed methods study results are provided in subsequent sections. Qualitative analysis of the primary data is presented first followed by a quantitative analysis of the secondary data.

The results validated how the management of vulnerable populations, such as TAY, is multifaceted and has been operating across various fragmented, siloed systems until the implementation of WCM in 2016. Qualitative analysis of primary data is presented first, followed by quantitative analysis of secondary data, highlighting how WCM is facilitating the streamlining of care processes and the resulting outcomes.

Primary Data: Qualitative Analysis

A total of 6 Aspiranet employees agreed to participate in the study. Table 1 summarizes the participants' characteristics. All were aged >30 years (median 37, IQR 32-39) and have either a bachelor's or master's degree, and the majority (5/6, 83%) were females. Participants included leaders (4/6, 67%) and life coaches (2/6, 33%) having a median of 6 (IQR 5-10) years of experience at Aspiranet and overall experience in this field of 10 (IQR 7-14) years. Regarding self-assessed technology skills, most (4/6, 67%) participants rated themselves as experts or as close to an expert. The median time spent on each interview was approximately 56 (IQR 53-75) minutes.

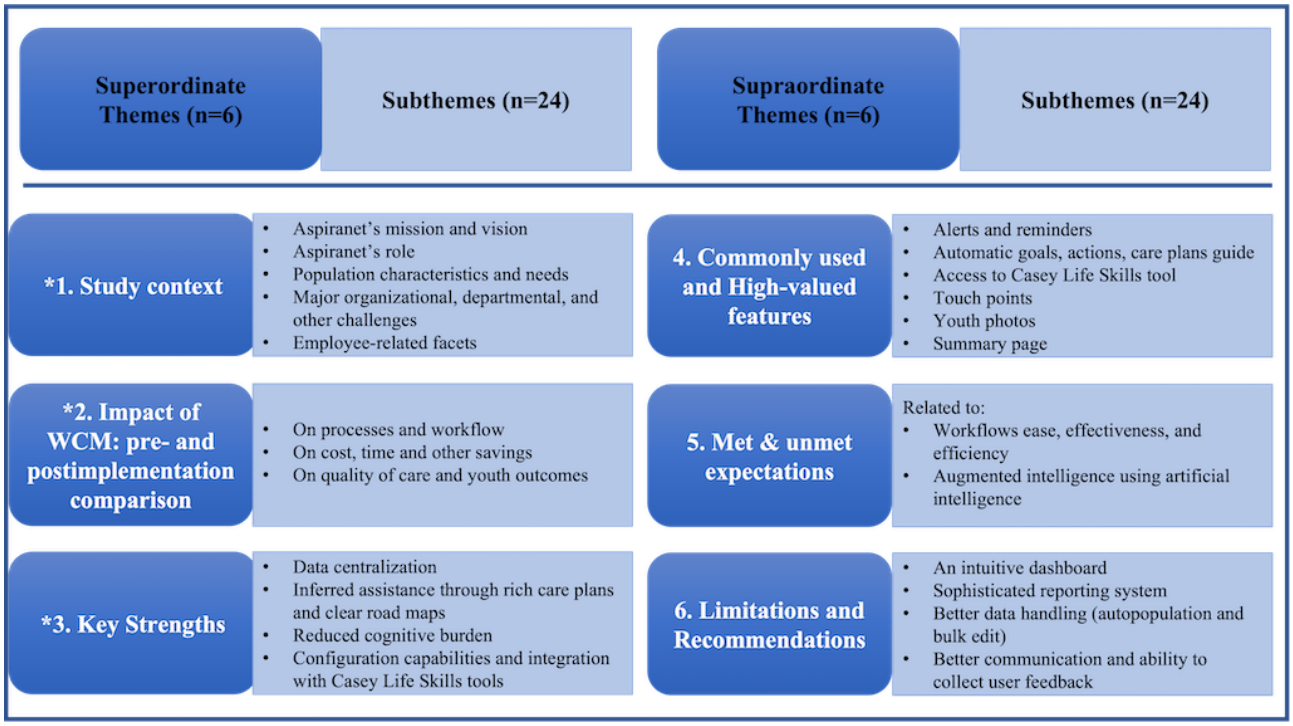
The thematic analysis of 6 transcripts by 2 qualitative experts (RFR and CBV) unveiled various subthemes (n=24) that were grouped into the following higher level superordinate themes (n=6), as summarized in Figure 1.

The 6 superordinate themes and the corresponding subthemes that emerged from the thematic analysis of the interview transcripts are described in subsequent sections.

Table 1. Study participant characteristics (n=6).

ID	Gender	Age (years)	Degree	Role	Experience at Aspiranet (years)	Overall Experience (years)	Self-assessed technology skill level	Interview duration (minutes)
1	Female	37	Master's	Life coach	5	7	Intermediate	52:52
2	Female	32	Bachelor's	Life coach	6	10	Expert	67:02
3	Female	31	Master's	Leader	3	5	Expert	48:03
4	Female	63	Bachelor's	Leader	10	10	Close to an expert	54:43
5	Male	39	Bachelor's	Leader	12	15	Between intermediate and expert	75:37
6	Female	36	Master's	Leader	6	14	Close to an expert	58:28

Figure 1. Thematic analysis reporting superordinate and subthemes (*key themes). WCM: Watson Care Manager.



Superordinate Theme: Study Context

This theme has subthemes that are explained in subsequent sections.

Aspiranet's Mission and Vision

The mission of Aspiranet's TAY program is to provide young adults between the ages of 18 and 24 years with a level of support and services so they can be successful once they transition out of foster care and start living on their own. The TAY program's long-term vision is to help support a world in which all children, youths, and their families are loved and cared for and that they have the resources that they need to lead a healthy and successful life.

This is a complex challenge. Aspiranet's life coaches help link their clients with governmental social resources, while also helping them develop the skills needed to achieve basic life skills (eg, housekeeping) and supporting their path to achieving and maintaining safety, permanency, and economic success (Textbox 1). In addition to the diversity of client demographics (eg, age, race, ethnicity, and religion), life coaches may help address a range of issues from emotional, behavioral, and mental health conditions to teen pregnancy, homelessness, and employment challenges. Life coaches' clients often require dramatically different levels and types of assistance.

Textbox 1. Aspiranet staff feedback on their mission and vision.

- Aspiranet's mission
  - "I think, (the role of) Aspiranet is to provide our clients with the foundation & support to be able to kind of live on their own & provide resources, (that could guide) them to transition from foster care to a more independent living." [Life coach]
- Aspiranet's vision
  - "But, I mean, our vision is a world in which all of our children, & for my division is youth are loved & cared for, you know, & that they have the resources that they need to try." [Leader]

Aspiranet's Role

The role of Aspiranet starts with first breaking down communication barriers with the youth who often come in with multiple traumatic experiences, rendering them emotionally unstable. This leads to an open and more productive discussion between the youth and caregivers and helps to identify the youth's specific needs (education, housing, employment, and basic life skills) and to make targeted plans. Aspiranet plays an important role to help youth achieve the actual goal by linking resources to needs through collaboration with various agencies (Textbox 2). To make changes at the macro level, Aspiranet plays an essential role in engaging with the legislative process, particularly at the state level:

Textbox 2. Aspiranet staff feedback on their role with at risk youth.

- "I was going to say, it's just kind of my job is kind of like the tour guide. It's like, I'm like that step. Yeah, like on their side so, like, as they're kind of moving along, I'm literally there through every step." [Life coach]

Population Characteristics and Needs

All interviewees mentioned the prevailing diversity in youth population demographics in terms of gender, age (18-24 years), race and ethnicity, and religious backgrounds. All patients experienced some type of trauma in their lives. They all pointed out an array of pressing challenges that their clients face (Textbox 3), often more than one at a time. Examples include age-related emotional challenges, behavioral and mental health conditions, housing, education, employment challenges, and lack of basic life skills as they are often systemized (ie, coming from structured facilities with more dependency on others), multiple placements and acquaintance to environment and individuals, and pregnancy at a young age:

Textbox 3. Aspiranet staff feedback on challenges that their incoming youth encountered.

- "I would say housing just because the market for this particular population is scrutinized sometimes only because they're young, so when it comes down to being 21, a lot of apartment complexes don't necessarily like to work with 21 year old. Landlords think our clients are not mature enough to handle the pressure of being on their own. So, it's a lot of the housing & on top of them having the housing crisis that we've had for a few years." [Leader]

Major Organizational, Departmental, and Data Location or Complexity Challenges

Several challenges were brought up by leaders and life coaches and grouped into 2 main categories: one independent of the COVID-19 pandemic and the other ensuing or surfacing during the COVID-19 pandemic. Agencies working in silos owing to lack of integration and coordination are another important challenge. The continuing reliance on paper and not being fully digitalized added more challenges to the processes. Some challenges noted include having large amounts of data from different data resources; data existing in various formats and in different locations; duplicate data entry; and not being able to comprehend the full picture of the youth's records from multiple discrete pieces of data. Financial challenges and limited funding, especially to meet housing needs in the state of California with the rising cost of living, were mentioned most often. Among other challenges that are directly or indirectly affected by the COVID-19 pandemic are those related to hiring staff, developing and training new employees, and excessive workloads. (Textbox 4)

Textbox 4. Aspiranet staff feedback on their day-to-day challenges.

- Lack of integration and coordination
  - "You know, the lack of integration or coordination. With other departments & agencies, we often work in silos where other providers are out there, providing the same type of work that we're providing & kind of doing it in silos." [Leader]
- Reliance on paper-based data
  - "...another issue is just in terms of the integration of services, it's really this way of working where we're still very reliant on paper..." [Leader]
- Fragmented data (in discrete formats and locations)
  - "...(we) collect a lot of data associated with both our youth as well as the services we're providing. Then (there is data) on an ongoing basis (such as) case notes, service plans, all sorts of things... for each youth that we serve &...lot of cases, those pieces of information are in lots of different locations or different systems." [Leader]
- Challenges associated with the pandemic
  - "Staff development, the workload being too high, insufficient time, specifically over the last year during the pandemic." [Life coach]

Employee-Related Facets

An employee's overall integrity and commitment to help clients achieve their goals is considered an indirect measure of their performance. Therefore, it is the quality and not the quantity of work that steers the performance evaluation (Textbox 5). All employees mentioned that the ongoing training they periodically receive helped them better understand the processes and better use available tools, particularly WCM:

Textbox 5. Aspiranet staff reflect on their performance related measures.

- "At least from what I've been able to see, and how I evaluate the employee, it is not really about the quantitative work, (it is) more about the richness of the work." [Leader]

Superordinate Theme: User-Reported Benefits of WCM

This theme has subthemes that are explained in subsequent sections.

Impact on Workflow Processes

Beyond the convenience of access and minimizing the gathering of paperwork, all respondents brought up the benefits of having a centralized electronic system in which data were entered and saved in real time. Rather than entering weekly contact notes (before the WCM), notes are being entered in a timely manner without waiting for weekly data dumps. Data could now be retrieved and viewed immediately from anywhere, anytime, and anyone from the team.

WCM enhanced the usability and utility of youth information through the ability of configuration capabilities to retrieve and present desired reports. The participants discussed the importance of effective and efficient workflows and how this efficiency enabled them to rapidly triage vulnerable residents to the services they need in a quicker manner. This ease of access to a full synopsis of any client allows for flexibility in the performance of client duties. In an emergency, a life coach unfamiliar with a client could simply look up all their pertinent information. Although not a perfect solution, this capability did provide a significant improvement over the current workflow (Textbox 6).

Textbox 6. Aspiranet staff reflect on the transition before and after implementation of Watson Care Manager.

- Real leader time electronic data entry
  - "Prior to WCM, (the life coaches) had to write down notes then go to the file cabinet to pull out notes or to put them into a file & so now with Watson, the notes get entered in real time... as they're having their interaction with the young adult, they're already typing in the notes." [Leader]
- Effective or timely care delivery
  - "...(the process is) transitioning...to a real time information mindset. In other words, the data isn't updated once a week, or once a quarter. It is updated once a day; it's updated as frequently as it needs to be...Personally, I think that's a huge shift in the way people work & I think it's a huge shift in the way people are able to kind of better manage their youth relationship."
- Easy access to data
  - "Before, you know, we just had it (paper records) in a chart room. You had to be literally in the office to find anything you wanted. So, it's really great to be able to have that on hand, made it smoother & when, in emergencies or in crisis where you need to access that information quickly." [Life coach]

Impact on Time and Cost Savings

Participants believed that there were definite time savings realized after WCM implementation that indirectly or directly resulted in cost savings. Some of the reasons included saving time on commute due to convenient access to the tool, less overtime granted to catch up on notes, and more effective notes retrieval (Textbox 7).

Textbox 7. Aspiranet staff identify how Watson Care Manager potentially affected time savings.

- Time savings
  - "Case so absolutely, I would say a savings in the travel & the time in printing & scanning & in having to set up meetings because now it's more collaborative & people are more up to speed or the cases." [Leader]
- Cost saving as proxy of time savings
  - "Because a lot of overtime was granted catching up on notes... there's been tons of time savings... Watson is available on electronic devices. It's mobile. They don't have to come back to an office to write down their notes or to type things, & they're doing it out in the field. I'm 100% positive that it has saved our life coaches a lot of time." [Leader]

Impact on Quality of Care and Outcomes

According to the participants, their interactions with youth are much more enriched; that is, youths are more focused and engaged while managing their own care plans. There is consensus that youth outcomes are also improving, such as better retention, which could be further validated by examining objective data. However, the reason for these improvements may be multifactorial because WCM is only one of several modifications to workflows (Textbox 8).

Textbox 8. Aspiranet staff discuss how Watson Care Manager impacted their work with the youth.

- Enriched interactions and better engagement
  - "Definitely, I could see (improvement), like, it kind of helps to enhance in a sense because, like I said, the youth can more physically see what's going on...so that does help because it's (they) more focused..." [Life coach]
  - "...but I think definitely (there is an improvement)...but from my observation (there is) an increase in youth engagement." [Leader]
- Better outcomes
  - "...[There is]decrease in voluntary & involuntary discharges. Some of the data that I'm seeing [outcomes] right now can I tie it to Watson 100% confidently? No! but I think that there's a correlation." [Leader]

Superordinate Theme: Key Strengths

This theme has 4 subthemes explained in subsequent sections with supporting quotes.

Data Centralization

The tool served as a convenient, centralized system, with all youth information consolidated in one place, thus providing a holistic 360° view of any youth at any time (Textbox 9).

Textbox 9. Aspiranet staff discuss how Watson Care Manager consolidated their workflow.

- Centralized data hub
  - "Oh, yeah, having Watson Care Manager keeps everything in the centralized system. That way we don't have to go back and forth in between pulling up the youth's goals in their folder. We have it all in our Watson Care Manager." [Leader]
- Holistic view of youth profile
  - "...here again, this kind of a 360 degree view that's accessible to any member of a care team at any point." [Leader]

Inferred Assistance Through Rich Care Plans and Clear Road Maps

WCM helped all youth receive high-quality care, regardless of the complexity of the youth's needs or the experience level of the life coach by rendering rich care plans with clear road maps. As stated earlier, many life coaches had heavy client burdens, and leaders found it difficult to hire new staff in the wake of the COVID-19 pandemic. As such, life coaches had large client loads, and a number of coaches had few years of experience. The large client list and complexity of each client case meant that each client required an array of assessments, documents, and other appraisals to ensure their health and safety (Textbox 10). By embedding access to decision support, WCM supported life coaches with varying levels of experience to perform treatment protocols developed by seasoned veterans:

Textbox 10. Aspiranet staff reflect on how Watson Care Manager assisted them in complex tasks.

- "Say, if I'm putting in a safety plan, it will actually alert me to ask 'did I contact any medical needs? Are the police involved?' It (WCM)reminds me 'hey, you need to check on these things'... So, you're not like trying to go off of experience... I think really guides our best practice." [Life coach]

Reduced Cognitive Burden

In this capacity, WCM functioned as a "memory bank," ensuring that no tasks or actions needing attention are overlooked or missed action (Textbox 11). This is provisioned through the WCM capability of alerts and prompts, in addition to real-time data entry, eliminating any recall biases while entering data:

Textbox 11. Aspiranet staff reflect that Watson Care Manager would aid them in the full completion of tasks.

- "Watson will project for specific actions that need to be done to that task. So, say, for instance, if I put in a safety plan, it's going to prompt me (and ask) have you done this, this, this and this? So instead of me trying to operate on memory, Watson's actually projecting and saying, 'stop! -do this right now or else'." [Life coach]

Configuration Capabilities and Integration With CLS

The integration of WCM with the CLS tool has provided us with an opportunity to follow youth progress in a more intuitive and efficient manner. Earlier, the role of the CLS tool was more for compliance checks but now it is being used more meaningfully (Textbox 12):

Textbox 12. Aspiranet staff feedback on the intuitive integration of Casey Life Skills into their workflow.

- "I am very excited about the configuration within Watson to our Casey life skills tool. Traditionally, we use this tool and (it was) just filed away... it was more transactional compliance check, completed it every 6 months, and it was thrown into the file. (But) With the configuration that we did within Watson recently, we were able (have) the scores added to the thing in Watson and then we're able to pull a report from Cognos to be able to demonstrate growth." [Leader]

Superordinate Theme: Commonly Used or High-Valued Features

The users mentioned several commonly used features. A few that were brought up repeatedly are alerts and reminders, automatic goals or action, configurability with CLS, touchpoint notes, youth photos, and the summary page (Textbox 13).

Textbox 13. Aspiranet staff feedback on their most commonly used Watson Care Manager features in their workflows.

- Alerts and reminders
  - "Watson so that we get alerted when the system recognizes. Someone may benefit for mental health services so that the life coach, and the team is aware, and can provide those referrals to the community to address any gaps." [Leader]
- Automatic goals, actions, and care plans
  - "So, definitely the automatic goals action alerts. I think really guide our best practice and are used the most often." [Leader]
- Configurability with Casey Life Skills
  - "Casey's (CLS) are actually interacting with the goals so that's really nice. That's a new feature that's just kind of been developing. And so that's really nice. I do like that. It promotes tasks. Like I said, I enjoy that too." [Life coach]

Superordinate Theme: Met and Unmet Expectations

Workflow Ease, Effectiveness, and Efficiency

One of the initial expectations was that the new care management tool would make the completion of the processes and associated tasks more efficient and effective. Employees' expectations were met by the tool's implementation, as it brings all data into an electronic form and under one platform (Textbox 14). Previously, information was collected and stored in various places and different formats:

Textbox 14. Aspiranet staff reflect on their expectations moving to an all-in-one platform.

- "One of the expectations (we had and) was not an immediate expectation. Something that would happen over time is that there would be kind of efficiencies associated with the work that a life coach did and improvement in their process... Being able to, you know, access data and enter data in the field to have data all in one place, as opposed to in a lot of different (place), you know, systems or paper files or whatever. And I think to that extent, Watson has met those expectations." [Leader]

Augmented Intelligence Using Artificial Intelligence

Artificial intelligence technologies was an area where employees had higher expectations that they felt remained unfulfilled (Textbox 15). They believed that since the technology is still evolving, their expectations would be met in the near future:

Textbox 15. Aspiranet staff reflect on their expectation of an artificial intelligence technology.

- "You know, I was expecting a robot to be honest almost. Maybe, I was not being realistic that (it) was going to take over (and) is going to help?? (us) through every single step...I had very high expectations of it and maybe unfair expectations of it, you know so I would say that while it has not fully met my entire expectation of it. You know, it's making steps towards that very slowly..." [Leader]

Superordinate Theme: Limitations and Recommendations

Intuitive Client Dashboard

Having a client dashboard for users and to generate reports enabling them to see their client snapshots is another feature currently absent that would improve the workflow (Textbox 16):

Textbox 16. Aspiranet staff reflect on what potential features like dashboards that would be helpful in the future.

- "And so, like I said, it, that part has been hard to like (i.e.), to run a quick report...also our supervisors can't (run reports), there is not like a quick dashboard. They would have to understand tables in order to go into Cognos and run the report to get that information." [Leader]

Enhanced Information Retrieval Functionalities

Integration of a sophisticated search engine to further expedite information retrieval was a commonly reported theme (Textbox 17):

Textbox 17. Aspiranet staff reflect on what potential features like search engines that would be helpful in the future.

- "If you can search, like, a search window where you can type (such as), if you are looking for some specific doctor's visit, so type in (doctor's visit), and it pulls up all the doctor's visit. Something like that." [Life coach]

Efficient Data Management

Data management, such as automating the data population, the ability to bulk edit, and repeated data entry tasks, is believed to be suboptimal. Additional developments around these features could likely reduce the redundancy of task completion and hence enhance efficiency (Textbox 18):

Textbox 18. Aspiranet staff reflect on efficiencies that could be gained through data automation.

- "(Getting tasks done) is tedious (having), repetitive forms or repetitive actions and tasks or just having their name and birthday (repeatedly entered) in every form, it just gets to be too much..." [Life coach]

Enhanced Communications Functionalities

Some of the recommendations made by the employees were enhancing the communications functionality within the tool; for example, provision of email or chat between supervisors and life coaches for quick communications (Textbox 19):

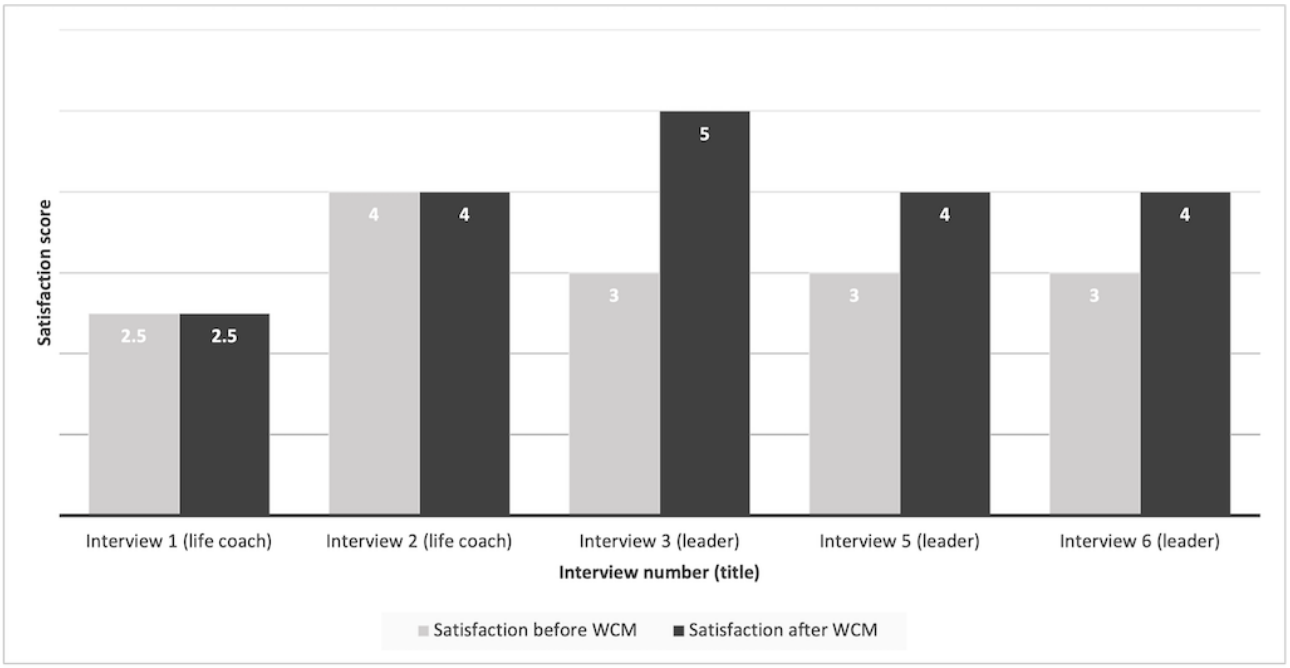
Textbox 19. Aspiranet staff reflect on how their recommendations for WCM would improve internal communication.

- "Supervisor needs to review something that a life coach completed, there should be like a way to alert them through the tool. Or I feel like, in this case, it would be an email that needs to be sent saying, 'hey, you have an alert on this client'." [Life coach]

Primary Data: Quantitative Analysis

When comparing pre- and post-WCM implementation satisfaction scores, as measured using a 5-point Likert scale, the scores remained constant between the 2 life coaches (median 3.25, IQR 2.5-4). Among leaders, the results suggest a higher satisfaction score for post-WCM implementation (median 4, IQR 4-5) when compared with pre-WCM implementation (median 3, IQR 3-3; Figure 2). Data from interview 4 were not included because there was no before and after comparison.

Figure 2. Pre– and post–Watson Care Manager implementation satisfaction scores measured on a 5-point Likert scale. The number denotes self-scored satisfaction.



Secondary Data: Quantitative Analysis

Youth Demographics, Length of Service Days, School Enrollment, and Employment

Across the 2 study phases, the Aspiranet TAY program served 1641 transitional youth. The median age (IQR) was 18 (18-19) years, and 55.32% (614/1641) of the patients were female. The reported race or ethnicity data across this population included Hispanic or Latino (621/1641, 37.84%), Black (470/1641, 28.64%), White (397/1641, 24.19%), and other (153/1641, 9.32%; reported in Table 2).

Table 2. Youth demographics and social determinants of health characteristics pre– and post–Watson Care Manager (WCM) implementation.

Youth characteristic	Pre-WCM (n=531)	Post-WCM (n=1110)
Age at start (years), median (IQR)	18 (18-19)	18 (18-19)
Gender (female), n (%)	289 (54.4)	614 (55.32)
Race and ethnicity, n (%)		
Hispanic or Latino	202 (38)	419 (37.75)
Black	152 (28.6)	318 (28.65)
White	126 (23.7)	271 (24.41)
Other	51 (9.6)	102 (9.19)
Length of service (days), median (IQR)	234 (130-479)	249 (123-458)
Current school status, n (%)		
Enrolled full time	359 (67.6)	833 (75.05)
Enrolled part time	61 (11.5)	91 (8.2)
Not enrolled	111 (20.9)	186 (16.76)
Employed, n (%)	137 (25.8)	152 (13.69)
Housing status, n (%)		
Vacancy (existing)	130 (24.5)	885 (79.73)

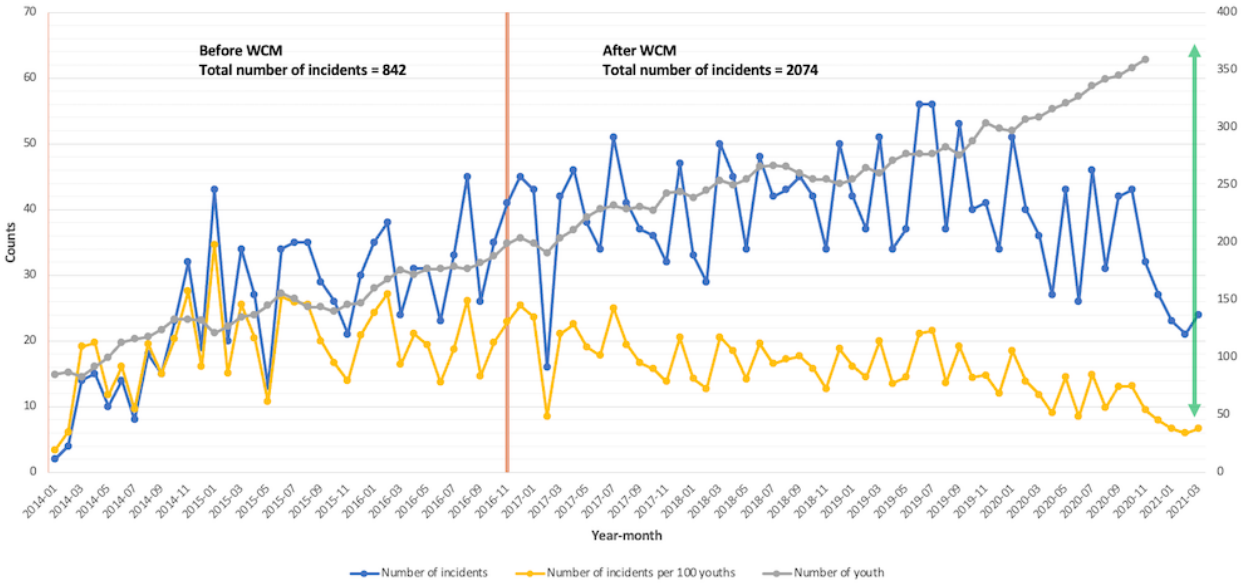
The overall median length of service (IQR; ie, time spent under care management) was 247 (125-468). School enrollment varied, with most youth enrolled full-time (1192/1641, 72.64%), some enrolled part time (152/1641, 9.26%), or some not enrolled at all (297/1641, 18.1%). Youth employment and available housing vacancy were 17.61% (289/1641) and 61.85% (1015/1641), respectively.

Table 2 summarizes the characteristics of the pre- and post-WCM implementation. The data showed that youth demographics and median length of service days remained consistent for both phases. The school status showed some increase in full-time enrollment, a decrease in part-time enrollment, and a reduction in the number of students not enrolled. The percentage of youths employed from the pre-WCM to post-WCM phase declined, likely impacted by the COVID-19 pandemic. With respect to housing, the number of available vacancies increased substantially from before to after the WCM.

Incidence of Undesired Events Among Youths

For the 2 phases under study, the number of incidents (undesired events) reported was 842 and 2074 for pre- and post-WCM implementation, respectively. Figure 3 shows the monthly trend for the number of incidents (blue line), number of youths receiving services (gray line), and number of incidents per 100 youths (yellow line). The number of youths receiving services increased over time (gray), while the number of incidents (blue) seemed to increase slightly over time, even though variability was high. The results also suggest that, although there is variability, the number of incidents per 100 youths (yellow) shows a slight decrease over time.

Figure 3. Number of incidents, number of youths receiving services, and number of incidents per 100 youths. The red line delineates the before and after phases, whereas the green arrow indicates the widening gap between the number of youths (increasing) and the number of incidents (decreasing). WCM: Watson Care Manager.



Discussion

Principal Findings

Providing extended foster care to help prepare disadvantaged young adults for transition into adulthood is a complex process. TAYs are often confronted with multiple challenges including social, emotional, physical, and mental instabilities (eg, lacking basic life skills, housing insecurities, criminal involvement, and depression or anxiety) simultaneously. Existing processes around managing those processes are often fragmented and resource intensive, which can interfere with effective and efficient care delivery and desired youth outcomes. It is imperative for organizations to take actions to overcome such roadblocks (eg, managing large amounts of typically paper-based data stored in various formats and at disjoint locations or building bridges to integrate existing silos in operational activities between disparate organizations involved in rendering services). The implementation of WCM at Aspiranet to assist in managing the care of youth transitioning from foster care to adulthood has enriched workflow processes and is positively impacting youth outcomes by the transformation from a traditional paper-based system to an integrated and interdisciplinary technology-based system. Using WCM, Aspiranet's leaders and life coaches can develop integrated care plans for youth under their management. WCM helps in successfully triaging the vulnerable youth population to the optimum services they need in a more effective and efficient manner compared with the former highly manual, paper-based processes. In terms of youth outcomes, the results showed an increase in full-time school enrollment and a reduction in part-time school enrollment. A greater availability of vacant residential options serves as a proxy for favorable youth behaviors. Landlords' satisfaction is higher and their need has decreased in refuting TAY for housing privileges because of their hostile behaviors. The median number of days spent in the foster care program remained the same; however, the number of incidents reported per month per 100 youths showed a steady decline, even with an exponentially increasing number of enrolled youths (Figure 3).

Analysis of the interview data revealed that the implementation of the tool facilitated streamlining the workflow and efficient retrieval of youth information. The data are accessible anywhere and at any time, in contrast to previous methodologies that require digging through file cabinets and driving miles to access data. Currently, with WCM, information is more accessible and thus more usable for life coaches. Users reported more meaningful and productive interactions with the youth under their management, and the availability of a visual representation of the youths' progress and accomplishments enhances monitoring. Visualization helps develop hope and motivation among otherwise discouraged youth. Finally, quick access and sharing of information allowed users to respond to emergency situations faster, lowering the risk of undesired outcomes.

As observed in previous studies, we found that when new technology is introduced, there is often a disconnect between leaders and staff as they see things differently [32]. The same applies to the perception of WCM; that is, more favorable responses from leaders than from life coaches and direct users (with more frequent daily use) were identified in a recent study [19]. However, owing to the small sample size, we were unable to make any definite conclusions. Some plausible reasons may be that users showed frustration with repetitive data entry into the tool; the process was not fully electronic and the need for continued paperwork comprised approximately 20% to 30% of the total work; or usability-related challenges (eg, too many tabs that they are not familiar with, lacking functions such as limited formatting options, or missing bulk editing), lack of a dashboard, and a suboptimal reporting system. These learnings can be leveraged and used as opportunities for system improvement.

In terms of youth outcomes, full-time school enrollment increased, whereas part-time attendance declined, indicating that youth were more focused and able to successfully compete their goals. Employment decreased, possibly because of the COVID-19 pandemic, and the state of California revised school attendance laws during the pandemic. No change was observed in the length of service (ie, time spent under care management). The data showed a significant increase in available vacant housing, which could be attributed to WCM. The explanation is that WCM provided a better picture and understanding of the maintenance needs of the apartment complex or residential units, which resulted in a more proactive response to apartment landlord needs and sustained a strong relationship. According to one of the leaders, retaining the existing landlord or apartment relationship is a key factor that allows the program to continue to grow. Without available apartments for placement, youths in the TAY program would not exist, and the foster care program would not be as successful as it is today.

As the number of youths increased, the total number of undesired events such as fights, break-ins, falling grades, and police custody increased, but the monthly number of incidences showed a steady decline. This is promising and indicative of potential improvements in behaviors or actions.

With respect to the interpretation of the abovementioned data, we need to keep in mind that these outcomes could be influenced by a number of other variables. Further analysis of quantitative data is needed to make any claims that the benefits surface from the use of WCM per se. The earnings from this study could be used as a springboard to further enhance future artificial intelligence–based technologies in social care management. This could be achieved by improving and fortifying the tool's current strengths, identifying gaps from a larger group of diverse stakeholders, and addressing them to ensure that they align with their needs and improve youth outcomes.

Limitations

The study limitations include data challenges, specifically missing data and incomplete or even absent preimplementation use and outcome comparison data. This further illustrates WCM capabilities for storing and saving data in a more readily accessible and secure manner in a single place that is not previously available. Before WCM implementation, the data were recorded using pencil- and paper-based notes, spreadsheets, and databases with fragmented and missing information. A second limitation was a small interview sample size, especially with life coaches. A third limitation was the inability to interview the youth (vulnerable population) to gain an understanding of their perspectives.

Most importantly, the improvement seen in workflow processes or youth outcomes cannot be fully attributed to the implementation of WCM because there are likely multiple confounders, both known (eg, life coach to client ratio, workload, changes in process and service delivery road maps, and availability of resources including funding) or unknown, which require further investigation.

Conclusions

Foster care and social service agencies providing support to TAY are challenged by synthesizing large volumes of siloed data (such as health, social, behavioral, and judicial data) to support high-impact decisions. A care coordination approach that is supported by technology has many potential benefits. These include serving as a common, shared platform to consolidate and manage electronic data and to create and track specific goals and corresponding actions. These complex processes are now executed more effectively, efficiency gains are realized in real time, with greater perceived acceptability to end users, and ultimately positive outcomes for these youths to become independent and productive members of society.

Acknowledgments

The authors are thankful to the Aspiranet team (Anne Gables, Sharlene Vega, and Luis Madrid) and the IBM offering and client management team (Elizabeth Schnitter, Ali Hanif, Katie Keating, and Carol Hamilton) for their tremendous help and guidance with primary and secondary data collection and higher levels of interpretation of the study results.

Conflicts of Interest

RFR, CBV, VCW, BLR, BRS, MSL, DB, JL, JLS, and GPJ were employees of IBM at the time the study was conducted.

Multimedia Appendix 1
Interview transcripts used for the study.
[DOCX File , 24 KB - humanfactors_v9i4e39646_app1.docx]

References

1. Gümüscü A, Nygren L, Khoo E. Social work and the management of complexity in Swedish child welfare services. Nordic Soc Work Res 2020;10(3):257-269. [doi: 10.1080/2156857X.2018.1542336]
2. Fish S, Hardy M. Complex issues, complex solutions: applying complexity theory in social work practice. Nordic Soc Work Res 2015 Aug 26;5(sup1):98-114. [doi: 10.1080/2156857X.2015.1065902]
3. Service Array. Child Welfare Information Gateway. URL: https://www.childwelfare.gov/topics/systemwide/service-array/ [accessed 2021-04-21]

4. Courtney ME. The difficult transition to adulthood for foster youth in the US: implications for the state as corporate parent and commentaries. Soc Policy Report 2009 Mar 01;23(1):1-20. [doi: 10.1002/j.2379-3988.2009.tb00058.x]
5. AACAP Home. American Academy of Child and Adolescent Psychiatry. 2022 Jun 15. URL: https://www.aacap.org/ [accessed 2022-12-05]
6. Dworsky A, Courtney M. Addressing the mental health service needs of foster youth during the transition to adulthood: how big is the problem and what can states do? J Adolesc Health 2009 Jan;44(1):1-2. [doi: 10.1016/j.jadohealth.2008.10.134] [Medline: 19101450]
7. George RM. Employment outcomes for youth aging out of foster care. US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. 2002 Feb 28. URL: https://aspe.hhs.gov/reports/employment-outcomes-youth-aging-out-foster-care [accessed 2021-04-18]
8. Altschuler D, Stangler G, Berkley K, Burton L. Supporting Youth in Transition to Adulthood: Lessons Learned from Child Welfare and Juvenile Justice. Jim Casey Youth Opportunities Initiative. 2009 Apr. URL: https://assets.aecf.org/m/resourcedoc/jcyoi-supportingyouthintransitiontoadulthood-2009.pdf [accessed 2021-04-21]
9. Dworsky A, Napolitano L, Courtney M. Homelessness during the transition from foster care to adulthood. Am J Public Health 2013 Dec;103 Suppl 2:S318-S323. [doi: 10.2105/AJPH.2013.301455] [Medline: 24148065]
10. Oshima KM, Narendorf SC, McMillen JC. Pregnancy risk among older youth transitioning out of foster care. Child Youth Serv Rev 2013 Oct;35(10):1760-1765 [FREE Full text] [doi: 10.1016/j.childyouth.2013.08.001] [Medline: 24489422]
11. Extended Foster Care (AB 12). California Department of Social Services. URL: https://www.cdss.ca.gov/inforesources/foster-care/extended-foster-care-ab-12 [accessed 2021-04-18]
12. Rosenberg R, Abbott S. Supporting Older Youth Beyond Age 18: Examining Data and Trends in Extended Foster Care. Child Trends. 2019 Jun 3. URL: https://www.childtrends.org/publications/supporting-older-youth-beyond-age-18-examining-data-and-trends-in-extended-foster-care [accessed 2021-04-25]
13. Adoption and Foster Care Statistics. Children's Bureau. URL: https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/afcars [accessed 2021-04-25]
14. Williams SC. State-level Data for Understanding Child Welfare in the United States. Child Trends. 2021. URL: https://www.childtrends.org/publications/state-level-data-for-understanding-child-welfare-in-the-united-states [accessed 2021-04-25]
15. The California Youth Transitions to Adulthood Study (CalYOUTH). Chapin Hall at the University of Chicago. 2021. URL: https://www.chapinhall.org/research/calyouth/ [accessed 2021-04-25]
16. Feng H, Courtney ME, Okpych NJ. Memo from CalYOUTH. Chapin Hall at the University of Chicago. 2020 May. URL: https://www.chapinhall.org/wp-content/uploads/PDF/CY_YC_IB0520.pdf [accessed 2021-04-25]
17. Courtney ME, Charles P. Mental health and substance use problems and service utilization by transition-age foster youth: early findings from CalYOUTH. Chapin Hall at the University of Chicago. 2015. URL: https://www.chapinhall.org/wp-content/uploads/CY_MH_DP0515.pdf [accessed 2021-04-24]
18. Park K, Courtney ME, Okpych NJ, Nadon M. Chapin Hall Research Brief. Chapin Hall at the University of Chicago. 2020 Sep. URL: https://www.chapinhall.org/wp-content/uploads/CJ-involvement-memo.pdf [accessed 2021-04-18]
19. Rizvi R, VanHouten C, Bright TJ, McKillop MM, Alevy S, Brotman D, et al. The perceived impact and usability of a care management and coordination system in delivering services to vulnerable populations: mixed methods study. J Med Internet Res 2021 Mar 12;23(3):e24122 [FREE Full text] [doi: 10.2196/24122] [Medline: 33709928]
20. Silos to Systems: Preserving and Strengthening Families and Children Experiencing Recurring Child Welfare System Encounters and Housing Crises. The Corporation for Supportive Housing. 2011 Oct 6. URL: http://www.csh.org/wp-content/uploads/2012/02/SilostoSystemsChildWelfare.pdf [accessed 2021-04-19]
21. Johnson D, Chung P, Schroeder J, Meyers J. Bridging silos, improving systems. Foundation Rev 2012 Jun 01;4(2):84-97. [doi: 10.4087/FOUNDATIONREVIEW-D-11-00020.1]
22. McKillop M, Snowdon J, Willis VC, Alevy S, Rizvi R, Rewalt K, et al. A system to support diverse social program management. JMIR Med Inform 2021 Aug 30;9(8):e23219 [FREE Full text] [doi: 10.2196/23219] [Medline: 34459741]
23. Adoption and Foster Care in California. Aspiranet Child and Family Services. URL: https://www.aspiranet.org [accessed 2021-04-18]
24. Transitional Aged Youth Programs in California - Foster Care for Teens. Aspiranet. URL: https://www.aspiranet.org/transitional-aged-youth [accessed 2021-04-18]
25. California Adoption Agency and Foster Care Organization. Aspiranet. URL: https://www.aspiranet.org/about [accessed 2021-04-18]
26. IBM Watson Care Manager. IBM Corp. 2021 Mar 10. URL: https://www.ibm.com/products/watson-care-manager [accessed 2021-04-18]
27. Casey Life Skills Toolkit. Casey Family Programs. 2021 Nov 8. URL: https://www.casey.org/casey-life-skills/#:~:text=Casey%20Life%20Skills%20(CLS)%20is,living%20and%20self%2Dcare%20activities [accessed 2021-04-25]
28. NVivo. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-04-25]
29. Chapman AL, Hadfield M, Chapman CJ. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. J R Coll Physicians Edinb 2015;45(3):201-205. [doi: 10.4997/JRCPE.2015.305] [Medline: 26517098]

30. Williams M, Moser T. The art of coding and thematic exploration in qualitative research. Int Manag Rev 2019;15(1):45-55.
31. World Programming. URL: https://www.worldprogramming.com/ [accessed 2021-04-27]
32. Tech at work and the employee experience. PricewaterhouseCoopers. 2022. URL: https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/tech-at-work.html [accessed 2022-06-14]

Abbreviations

CLS: Casey Life Skills
TAY: transitional-aged youth
WCM: Watson Care Manager

Edited by A Kushniruk; submitted 17.05.22; peer-reviewed by S Pesälä; comments to author 08.06.22; revised version received 29.09.22; accepted 12.10.22; published 16.12.22.

Please cite as:
Rizvi RF, VanHouten CB, Willis VC, Rosario BL, South BR, Sands-Lincoln M, Brotman D, Lenert J, Snowdon JL, Jackson GP
Understanding a Care Management System's Role in Influencing a Transitional-Aged Youth Program's Practice: Mixed Methods Study
JMIR Hum Factors 2022;9(4):e39646
URL: https://humanfactors.jmir.org/2022/4/e39646
doi:10.2196/39646
PMID:36525294

Original Paper

From Testers to Cocreators—the Value of and Approaches to Successful Patient Engagement in the Development of eHealth Solutions: Qualitative Expert Interview Study

Christine Jacob[1], PhD; Steven Bourke[2], MSc, MPhil; Sabina Heuss[3], PhD, Prof Dr

[1]FHNW - University of Applied Sciences Northwestern Switzerland, Windisch, Switzerland

[2]PersonalPulse GmbH, Basel, Switzerland

[3]FHNW - University of Applied Sciences Northwestern Switzerland, Olten, Switzerland

Corresponding Author:
Christine Jacob, PhD
FHNW - University of Applied Sciences Northwestern Switzerland
Bahnhofstrasse 6
Windisch, 5210
Switzerland
Phone: 41 562027700
Email: christine.k.jacob@gmail.com

Abstract

Background: Research has shown that patient engagement is most commonly done at the beginning of research or to test readily available prototypes and less commonly done in other phases such as the execution phases. Previous studies have reported that patients are usually assigned a consultative rather than a decision-making role in health service planning and evaluation.

Objective: This study had 2 objectives: to better understand the challenges and opportunities in the inclusion of patients in the development of eHealth technologies and ideas on how to overcome the identified gaps and to create a research-based end-to-end practical blueprint that can guide the relevant stakeholders to successfully engage patients as cocreators in all human-centered design phases rather than mere testers of preplanned prototypes.

Methods: Key informant interviews were conducted using in-depth semistructured interviews with 20 participants from 6 countries across Europe. This was followed by a focus group to validate the initial findings. Participants encompassed all the relevant stakeholder groups including patient experts, eHealth experts, health technology providers, clinicians, pharma executives, and health insurance experts.

Results: This study shows that engaging patients in eHealth development can help provide different types of value; namely, identifying unmet needs, better usability and desirability, better fit into the patient journey, better adoption and stickiness, better health outcomes, advocacy and trust, a sense of purpose, and better health equity and access. However, the participants agreed that patients are usually engaged too late in the development process, mostly assuming a sounding role in testing a ready-made prototype. The justification for these gaps in engagement is driven by some prominent barriers, notably compliance risks, patient-related factors, power dynamics, patient engagement as lip service, poor value perception, lack of resources, mistrust, and inflexibility. On the positive side, the participants also reflected on facilitators for better patient engagement; for instance, engaging through engagement partners, novel approaches such as the rise of professional patient experts, embedding patients in development teams, expectation management, and professional moderation services.

Conclusions: Overcoming the current gaps in patient engagement in eHealth development requires consolidated efforts from all stakeholders in a complex health care ecosystem. The shift toward more patient-driven eHealth development requires education and awareness; frameworks to monitor and evaluate the value of patient engagement; regulatory clarity and simplification; platforms to facilitate patient access and identification; patient incentivization, transparency, and trust; and a mindset shift toward value-based health care.

(JMIR Hum Factors 2022;9(4):e41481) doi:10.2196/41481

KEYWORDS

telemedicine; smartphone; mobile phone; electronic health record; public health practice; technology; perception; health education; mobile health; mHealth; telehealth; eHealth; patients; patient engagement; patient voice; patient empowerment

Introduction

Background

There is growing evidence that patients who are better informed and more engaged in their own care are more likely to be knowledgeable, stick to their treatment plans, and have a better quality of life [1]. Contributing to the research about patient engagement in health innovation, such as eHealth tools, supports the paradigm shift needed to normalize the patient's role beyond "subject" or "participant" to "partner and cocreator" to the development of more effective eHealth solutions and ultimately lead to better health outcomes [2]. The insights provided by the individuals who care for and live with the disease on a daily basis are invaluable when it comes to creating innovative solutions that will be accepted and address unmet needs. However, patient engagement is often considered complex [3], and the definition of the term is not always clear, implying that different stakeholders may have a different understanding and different expectations of what patient engagement is [4]. This study adopts a definition that encompasses the key attributes of patient engagement, "personalization, access, commitment, and therapeutic alliance," and defines it as "the desire and capability to actively choose to participate in care in a way uniquely appropriate to the individual, in cooperation with a health care provider or institution, for the purposes of maximizing outcomes or improving experiences of care" [5].

Given its intricacies, research has shown that patient engagement is most commonly done at the beginning of research or to test readily available prototypes and less commonly done in other phases such as the execution phases [6]. Previous studies report that patients usually assume a consultative rather than decision-making role in health service planning and evaluation [7]. Nonetheless, true patient empowerment necessitates a shift from a *patients as testers* mentality to *patients as equal partners and cocreators*, which can be achieved by involving them in every step of the human-centered design (HCD) process. This may help optimize resources [8] and facilitate the development of health tools that address patients' real unmet needs [9], using their experience and knowledge of their condition and not solely based on professionals' perceptions of patients' needs.

Different barriers, including a lack of insight into appropriate engagement methods, may be limiting patient involvement in eHealth innovations development activities [10]. More research is required to validate the expressed views among the different stakeholders in the health care ecosystem to establish effective methods for engaging patients [7]. This work builds on the efforts made in traditional clinical research, such as the road map developed by Geissler et al [11], with the aim of extending them to eHealth through a HCD approach, which is still in its infancy when it comes to systematic patient engagement in health care innovations [12-14].

Engaging users in the development of eHealth tools is a critical factor for their success, as it safeguards their usability and safety [15]; this is reflected in how institutions such as the Food and Drug Administration demand evidence of end-user engagement in health technology design when reviewing market presubmissions [16]. There is a greater need to assist patients in the daily management of their disease, for example, by helping them develop better adherence to treatments, resulting in better care outcomes. Digital tools strive to provide opportunities for this management; however, current findings demonstrate that most mobile health apps do not necessarily increase medication adherence [17]. Therefore, in recent years, there has been a growing body of research involving users in the development of health care technologies in what is called HCD.

The term "user-centered design" was first coined by Donald Norman in the 1980s to refer to a design philosophy that puts technology users at the center of the development process [18]. In 2010, the ISO 9241-210 extended the definition to also include other stakeholders beyond the direct users of the technology and referred to this new approach as HCD. The standard describes the key benefits of the HCD approach by explaining that "usable systems can provide a number of benefits, including improved productivity, enhanced user well-being, avoidance of stress, increased accessibility, and reduced risk of harm" [19]. These concepts are closely related to universal design, which aims to develop accessible technologies for all users regardless of their physical or cognitive capabilities [20], creating an inclusive design that takes into account the often overlooked patient populations that may be facing physical or cognitive challenges due to their health conditions.

In health care studies, the user-centered design ISO 9241-210 for HCD of interactive systems [19], the HCD IDEO Field Guide to Human-Centered Design [21], and the Hasso Plattner Institute School of Design Thinking [22] are among the most used frameworks for this purpose [13]. These frameworks have some overlaps, and when aggregated, they would cover five key phases: (1) specifying the context and evidence review, (2) defining user requirements and user research, (3) producing the design and testing the concept, (4) prototyping and testing against the initial requirements, and (5) delivering the solutions and usability testing. However, eHealth providers and developers are often faced with factors such as time pressure and rapid development life cycles that render the structured and iterative nature of HCD challenging to apply in practice [15].

Objectives

This study had 2 objectives: to better understand the challenges and opportunities in the inclusion of patients in the development of eHealth technologies and ideas on how to overcome the identified gaps and to create a research-based end-to-end practical blueprint that can guide the relevant stakeholders to successfully engage patients as cocreators in all HCD phases rather than mere testers of preplanned prototypes. The resulting blueprint aims to support the key stakeholders across the health

care ecosystem to systematically cocreate with patients and assist in developing eHealth solutions tailored for people in specific contexts and with specific needs. This allows for ethical designs that respect privacy and quality of life and reduce the chances of situations with a high risk of human error, leading to the creation of more relevant and safer tools that are more likely to be adopted by their intended users for better health outcomes.

Methods

Overview

In this study, we adopted a qualitative paradigm, which has become more common in research concerned with the assessment of health technologies as well as health services; this was reflected in the rising numbers of qualitative research published in medical journals [23]. One of the reasons behind the growing importance of qualitative methods in health care research is that they enable us to understand the complexities of today's health care ecosystem by touching on complex social aspects such as user attitudes and behaviors in ways that cannot be reached by quantitative methods [24].

Scope and Conceptual Framework

The World Health Organization defines eHealth as "the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including health care services, health surveillance, health literature, and health education, knowledge and research" [25]. This study focuses on patient-facing eHealth tools, including self-management tools and remote eHealth solutions, and excludes tools with no patient interface, such as those used within and between care providers (eg, health care provider videoconferences or electronic health record integration), or health data analytics systems used at the population level.

Human-centric design has been chosen as the conceptual framework because it places the people we are trying to serve at the center and offers them the space to become partners in eHealth innovation. It is an iterative approach in which the focus is on understanding the dynamics between stakeholders across the ecosystem and cocreating with them. The framework allows for a systematic investigation of the gaps and possible engagement opportunities for each step of the design process, rather than only the testing phase, as is commonly the case. This systematic approach enables a better understanding of the barriers to patient engagement in the phases where they are currently least involved and discusses opportunities for better engagement strategies that cover all design phases.

Sampling Strategy and Participant Recruitment

As in most qualitative studies, this research used purposive sampling with the objective of generating rich insights [26]. Potential participants were recruited based on their ability to provide rich and in-depth information about the research topic; they had to be individuals who have personal experience with the topic being studied so that they can articulate their real-life experiences [26,27]. The main selection criteria were that participants must belong to one of the key relevant stakeholder groups (ie, patient experts, patient organizations, eHealth providers and developers, clinicians, pharmaceutical experts, payers, and health tech researchers) and must have eHealth knowledge and experience to ensure a comprehensive view that takes the different perspectives into account when identifying the existing gaps and challenges and how to overcome the existing gaps and challenges with strategic patient engagement points and their realization to extend the benefit for all involved parties.

After shortlisting the participants of interest, as per the criteria explained earlier, the researchers contacted the key informants. To minimize potential selection bias, the researchers worked with the key informants to identify suitable participants in their network, a sampling technique called snowballing, where the researcher builds the sample through the network of other participants, in this case, the key informants [26]. As for the sample size, it is common in qualitative research aiming to identify patterns throughout data to recruit a sample somewhere between 15 and 30 interviews [26]; therefore, the researchers aimed to recruit enough participants, with the aim of reaching saturation, which is usually a signal that enough data have been collected, which is when new data do not generate new insights anymore [26,28-30].

Table 1 presents the demographics and characteristics of the sample. Several participants had multiple backgrounds; for example, some were pharmaceutical executives who worked in pharmaceutical companies and then moved to work for a big tech company. They combined both backgrounds and shared insights that capitalized on their experiences in both worlds. Participants categorized as eHealth experts have combined expertise in developing, conceptualizing, or testing eHealth tools. The participant, categorized as a health insurance expert, works for a health insurance company and has in-depth expertise in eHealth assessment and reimbursement criteria. Those categorized as health care professionals are participants with a clinical background. The participant categorized as a patient advocate is not a patient but rather an expert involved in patient organizations and actively working on patient engagement initiatives, such as assessment frameworks. Patient experts combine their disease knowledge and experience with eHealth relevant professional expertise and skills such as software development and user experience. Pharmaceutical executives are participants who work or have previously worked for a pharmaceutical company. Furthermore, those categorized as technology providers are participants who work in either an eHealth startup or a big tech company that focuses on eHealth.

Table 1. Sample demographics and characteristics (N=20).

Demographics and characteristics	Values, n (%)
Background (some participants had multiple backgrounds)	
eHealth experts	6 (30)
Health insurance experts	1 (5)
Health care professionals (clinicians)	3 (15)
Patient advocates	1 (5)
Patient experts	7 (35)
Pharmaceutical executives	5 (25)
Technology providers (big tech and startups)	7 (35)
Sex	
Female	11 (55)
Male	9 (45)
Location	
Belgium	1 (5)
Germany	2 (10)
Ireland	4 (20)
Italy	1 (5)
Switzerland	9 (45)
United Kingdom	3 (15)

Data Collection and Synthesis

Data were collected via in-depth, semistructured interviews conducted on the web. Data collection took place from March to May 2022, and a total of 20 participants located in 6 countries across Europe (Switzerland, Germany, the United Kingdom, Ireland, Belgium, and Italy) were interviewed. The median interview duration was 62 minutes, resulting in 291 pages of transcribed interview data for the 20 interviews.

The high-level research questions that helped guide the one-on-one interviews with the relevant stakeholders to create the blueprint are as follows:

- What is the value of engaging patients in the development of eHealth technologies?
- What are the barriers to and facilitators of patients' engagement in eHealth tools development?
- What are the gaps in the current patient engagement approaches (ie, the development phases in which they are least involved)? What can be novel approaches to patient involvement in cocreation to overcome the current gaps?

These research questions resulted in an interview guide composed of 12 questions and a maturity assessment survey that reflected the 5 stages of the human-centered approach to design. A copy of the interview guide is included in Multimedia Appendix 1, and a copy of the Survey Monkey form for maturity assessment is included in Multimedia Appendix 2.

Data coding began with a preliminary data extraction grid that included themes informed by previous research, the systematic steps in the HCD framework, and the research team's previous work on the topic. More themes were added as they emerged during the data analysis process. The thematic analysis by Braun and Clarke [28,29] was used to identify and extract themes addressed in the research questions. Computer-assisted qualitative data analysis software, Atlas.ti, was used for data coding. The first author (CJ) conducted the interviews and performed the initial analysis and coding. The second author (SB) reviewed the coding, and any cases of disagreement were discussed in conjunction with the last author (SH) and mutually agreed upon. The phases of thematic analysis are explained in detail in Multimedia Appendix 3. This process lasted from May to July 2022.

After generating the initial results, the researchers shared and discussed them in a web-based focus group with the same participants to ensure the validity and reliability of the findings and capture any potential additional insights that the participants may add [31]. The focus group was recorded and analyzed using the same method used for the in-depth interviews.

Role of the Researchers

Researchers play a fundamental role in qualitative research; this is because they are considered the instrument of the research, and accordingly, the analysis and findings are impacted by their approach and the way of evaluating and understanding things [26,31]. This does not mean that anything would be accepted in qualitative research, as some critiques say, but rather that the researchers "tell one story among many" that could be told about these specific data [26].

The practice partner, (SB), the founder of PersonalPulse, is delivering transformation in citizen-led health care innovation,

working together with relevant stakeholders in the health care ecosystem to cocreate health care solutions that are relevant, usable, and sustainable [32]. PersonalPulse is run by patients, is run for patients, and collaborates with a wide network of patient experts in diverse disease areas to give them a voice and empower them as equal partners in the creation of new health care solutions. The research partners (CJ and SH), from the University of Applied Sciences, Northwestern Switzerland, are both seasoned health care experts and researchers, with vast health care experience in hospitals, pharmaceutical companies, and health care technologies both from practice and research perspectives.

This background empowered the research team with a strong and wide network in health care and enabled them to access key informants in the area of eHealth. This helped them with access to participants and also fostered a relaxed and mutually beneficial dialogue between them and the key informants. To minimize the risk of researcher bias during the interviews, the interviewers refrained from stating their own views on the matters being discussed to minimize the likelihood of a directive discussion and to enable the participants to freely express their opinions [33].

Ethical Considerations

The Ethics Committee of Northwest and Central Switzerland determined that ethics approval was not needed for this study, according to the Federal Act on Research involving Human Beings, article 2, paragraph 1 (reference number Req-2022-00119). All participants were briefed about the research background and signed a consent form agreeing to participate.

Results

The Meaning and Value of Patient Engagement in the Development of eHealth Technologies

As a first step, we wanted to better understand how the expert participants define good patient engagement and the different types of value that it may generate. Figure 1 shows the themes that emerged as a response to these 2 key questions and their respective subthemes, reflecting the frequency of each theme (frequencies reflect the number of participants who mentioned that specific theme).

When asked about their definition of good patient engagement in eHealth development, many participants said that in its simplest form, it is about bringing patients' voices to the process (8/20, 40%), but most of them went a step further to explain that it is also about real cocreation and partnership (6/20, 30%), empowering patients to make a difference in their quality of life as a whole (6/20, 30%), engaging them in the whole process from beginning to end (6/20, 30%), and integrating them as equal partners in the development process (6/20, 30%).

Truly engaging patients as equal partners in eHealth development can bring different types of value. Most participants agreed that one of the most prominent values it can bring is the ability to identify unmet needs (16/20, 80%), followed by better usability and desirability of the tools (15/20, 75%), which resulted in better adoption and stickiness (14/20, 70%), and a more holistic view that enables a better fit into the overall patient journey (14/20, 70%). It also fosters trust and advocacy (6/20, 30%). The tools' better adoption and stickiness also imply better health outcomes because of adherence (6/20, 30%), which enables the least technically capable patients to still be able to use those tools, resulting in better health equity and access (3/20, 15%). It also gives a sense of purpose to the developing team as they can relate better to the patients' needs and pain points (3/20, 15%).

These key themes and subthemes, their frequencies, and sample quotes about the meaning and value of patient engagement are summarized in Table 2.

Figure 1. The meaning and value of patient engagement in eHealth development. QoL: quality of life.

Table 2. The meaning and value of patient engagement as expressed by the participants (N=20).

Theme	Sample quotes
The meaning of patient engagement	
Bring patients' voice to the process (n=8, 40%)	• "A much greater focus on getting the patient voice and really not even just the patient voice, but the diversity of the input and running everything by getting everything prototyped in a design way before we go into writing a line of code" [P13-HCP-TP] • "...you just need to have space and allow them to be in such a workshop and hear their voice and listen to them describing that" [P4-TP-Ph]
Cocreation and partnership (n=6, 30%)	• "Ideally, you co-develop things with them, not just you just get their perspective, and that's it because it's also at the same time, you're also changing their behavior" [P16-HCP-Ph] • "I think it's really about partnering, right? We're partnering at eye level with another expert, let's say, in what this person has gone through or is experiencing in their daily life" [P9-Ph]
Empower them to make a difference in their QoL[a] (n=6, 30%)	• "I think their involvement is always crucial because that should always be the overall objective, really improving patient life" [P12-DE] • "...really to empower patients and enable patients to look after their own health" [P13-HCP-TP]
Engage them from A to Z (n=6, 30%)	• "For me, patient engagement is, I think, about involving patients of every type at every level, at every point" [P15-TP] • "I don't see patient engagement as—I do my project and now I send you a questionnaire. And now I ask you, do you like it?—That's for me, not patient engagement. Patient engagement is having them all the way." [P19-PE-DE]
Integrate them as equal partners in health care (n=6, 30%)	• "Good patient engagement, it's putting the patient at the—give them a seat on the table at the same level as everyone else" [P19-PE-DE] • "I think about patient engagement, I consider it as the same level interaction between patient and the provider environment" [P20-PE-DE]
The value of patient engagement	
Identify unmet needs (n=16, 80%)	• "So it just brings the value of not just understanding what is needed, but understanding how it's needed, and when it's needed, and what could actually be used afterwards" [P19-PE-DE] • "you get really a feeling of is this really something the patient would use afterwards, or need in their life, in their daily living, and in their world" [P5-PE-DE]
Better usability and desirability (n=15, 75%)	• "The objective is to have the best solution, the most usable and effective solution." [P10-DE] • "If you can basically involve patients with different types of levels of understanding earlier on in the process, you're more likely to get a product that actually is tailored to everybody's familiarity." [P11-PE]
Better adoption (stickiness; n=14, 70%)	• "Ideally, you co-develop things with them, not just you just get their perspective, and that's it because it's also at the same time, you're also changing their behavior...it's more sustainable" [P16-HCP-Ph] • "The most important thing is that you are sure that there's an acceptance of what you're doing, that patients, in the end of the day, take your product, your concept, your whatever because they want to use it" [P18-HCP-DE]
Fit into patient journey (n=14, 70%)	• "In order to build solutions that solve real world problems for patients, then you need to use this really deep insight into the person behind the disease and in the context of their daily life. And honestly you can't do that without working very closely with the patient" [P6-PE] • "We need a patient journey that is way, way, way more easy than it used to be, way more rewarding, more nudging... So where can patients help in patient engagement? I think they really need to be there to describe the reality" [P8-TP-Ph]
Advocacy and trust (n=6, 30%)	• "If you have contributed to developing and you see that this has been developed by also patients like you, then you are more also prone to use it" [P2-PE-DE] • "This will help with adoption in the end, because now, yeah, we've created ambassadors, right? We don't only work on this project with the community, but still, everybody owns this now. All the co-creators own this solution and they're just waiting to share it with everyone" [P9-Ph]

Theme	Sample quotes
Better health outcomes (n=6, 30%)	• "Patient engagement is being able to sustain that change in behavior over time to get the clinical outcomes" [P13-HCP-TP] • "So I guess in terms of impact as well on people's life and really improving your health or at least the daily life and managing symptoms" [P2-PE-DE]
Better health equity and access (n=3, 15%)	• "They're able to augment the face to face with the digital platform, they're able to use it as an extender of care" [P13-HCP-TP]
Sense of purpose (n=3, 15%)	• "Having that engagement creates a more powerful purpose for the team" [P14-TP-Ph] • "The people developing or thinking of building a solution, if they feel somehow identified with the patient, there's an energy in the team which you have not seen before...it's really identifying with the goal to find solution for this problem" [P1-In]

[a]QoL: quality of life.

Barriers and Facilitators for Patient Engagement in the Development of eHealth Technologies

We then discussed the participants' experiences with the most prominent barriers to and facilitators of engaging patients in eHealth development. Figure 2 shows the themes that emerged as a response to these 2 questions and their respective subthemes, reflecting the frequency of each theme (frequencies reflect the number of participants who mentioned that specific theme).

Barriers to patient engagement in eHealth development revolved around 8 key themes: compliance and regulatory, patient-related factors, power dynamics in the health care sector, patient engagement as lip service or corporate social responsibility, value perception, resources, mistrust, and lack of flexibility.

Compliance was the most prominent barrier, with participants mentioning the complexity of regulatory processes as a key hurdle (18/20, 90%). However, some participants pointed out that this may also be partly a perception issue in that some stakeholders may perceive compliance as more complex than it really is (9/20, 45%). Compliantly compensating patients for their engagement was also perceived as a hindrance (9/20, 45%), and lack of process clarity was also raised as an issue, especially for smaller eHealth providers that may not have the resources or in-house knowledge about all regulatory processes (7/20, 35%).

Participants expressed that some patient-related factors may also make it difficult to engage patients in eHealth development. Specifically, not only are patient identification (13/20, 65%) and patient access (11/20, 55%) a key hurdle but also some health-related constraints may render it difficult for some patients to engage (4/20, 20%) or some patients' lack of the needed skills to engage efficiently (1/20, 5%).

Power dynamics in the health care sector may also hinder patient engagement. Patients not being seen as equal partners (9/20, 45%), conflict of interests among the stakeholders (7/20, 35%), the economic model (4/20, 20%), patients not being given a safe space to express their needs and pain points (4/20, 20%), and the lack of decision power in many cases (3/20, 15%) were the most prominent subthemes mentioned by the participants in this regard.

Other barriers included patient engagement being considered a marketing activity or lip service by some of the stakeholders (12/20, 60%), the lack of clarity on the value that patient engagement may bring (11/20, 55%), resource constraints (11/20, 55%), mistrust between patients and some other stakeholders (11/20, 55%), and sometimes a mere inflexibility of some eHealth providers (4/20, 20%).

These key themes and subthemes, their frequencies, and some sample quotes about barriers to patient engagement are summarized in Multimedia Appendix 4.

When asked about facilitators of patient engagement in eHealth development, most participants talked about working with different types of engagement partners to overcome some of the barriers such as compliance and patient identification and access. Patient organizations and advocacy groups were the most cited engagement partners (13/20, 65%), followed by clinicians (8/20, 40%), involvement in the patient community in general (6/20, 30%), and web-based patient communities (5/20, 25%).

Many participants also mentioned the rise of some novel approaches that play an active role in facilitating patient engagement, such as professional patient experts (13/20, 65%) and patient engagement agencies that play the role of matchmaking patients with the relevant stakeholders interested in engaging them (4/20, 20%).

Other approaches that may enable patient engagement include embedding them in the development team (12/20, 60%), managing patient expectations from the beginning to avoid disappointment in case some of their requests were not feasible or not in scope (7/20, 35%), and using professional moderation services that can help translate technical language to the patients and also help technical staff to understand the real needs of patients (6/20, 30%).

The key themes and subthemes, their frequencies, and sample quotes about the facilitators of patient engagement are summarized in Table 3 for clarity.

Figure 2. Barriers to and facilitators of patient engagement in eHealth development. CSR: corporate social responsibility; PE: patient engagement.

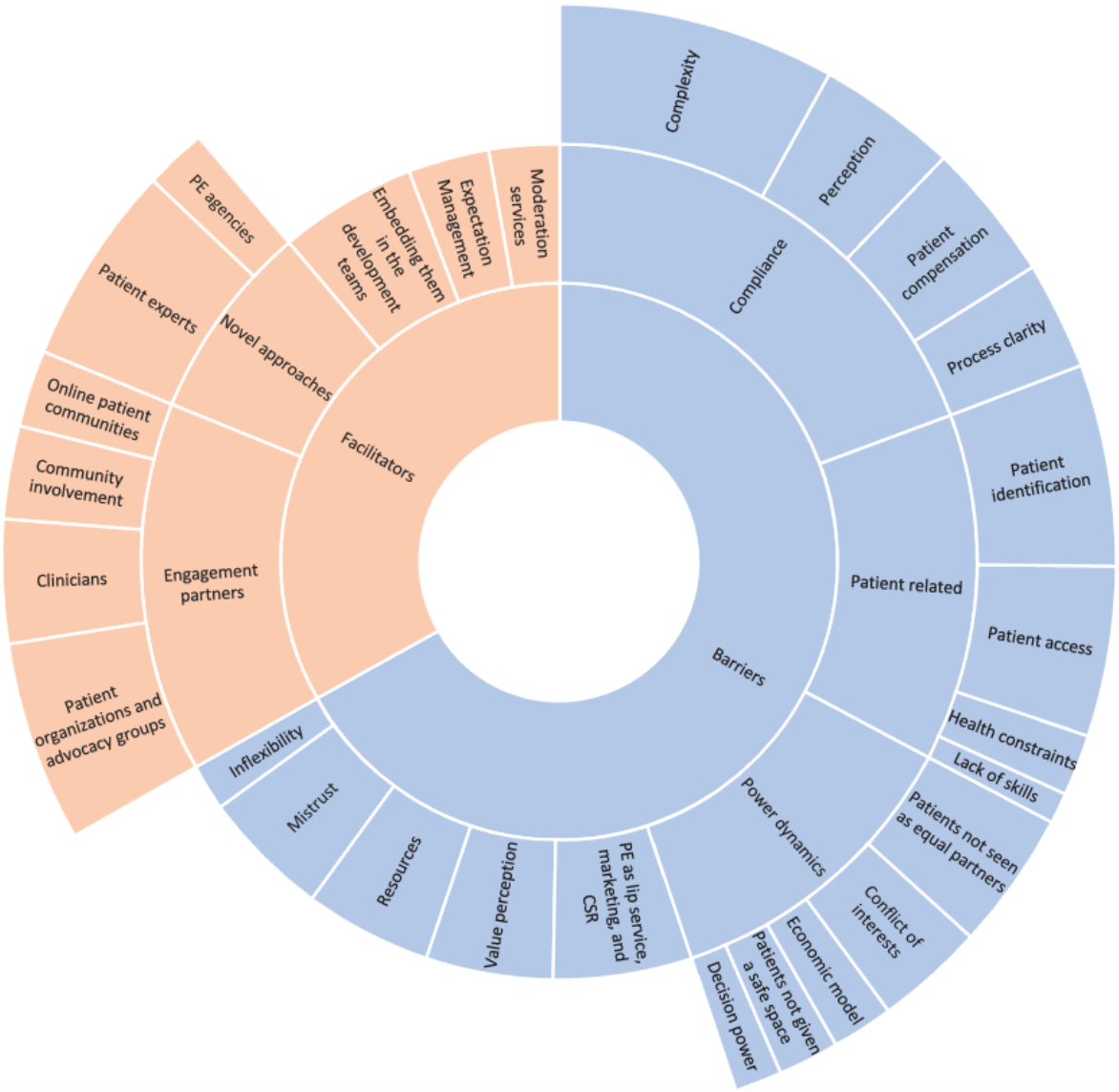


Table 3. Facilitators to patient engagement as expressed by the participants (N=20).

Theme	Sample quotes
Engagement partners	
Patient organizations and advocacy groups (n=13, 65%)	• "I also think that patient advocacy groups, of course, can support a lot because they can act as an important connector." [P12-DE] • "there's lots of good societies and charities that work with patients, you could approach them directly, try get involved with them." [P3-TP]
Clinicians (n=8, 40%)	• "The key thing there is it was the clinicians that had this trusted relationship with the patients and ask them to do this." [P13-HCP-TP] • "So, to find a patient through a specialist who knew which patient he could ask for and also knew how to bring them into a group and introduce them." [P1-In]
Community involvement (n=6, 30%)	• "I think being more involved in the community that you're serving as well. So as a company, taking part in charity events relating to the health area, maybe, that you're working in. Going along and volunteering at events, opening up opportunities." [P15-TP] • "And even being present in forums... how are we present in those discussions? Patients will find themselves in communities, social media, all of that. And that's a great way of scoping out new opportunities also. But somehow the organizations need to be present in those discussions." [P4-TP-Ph]
Web-based patient communities (n=5, 25%)	• "So, I think partnership with the patients, so whether it's a patient organization or the online patient community, but for me, I always say I think it needs to be a mixture because it needs to truly represent the community as a whole." [P16-HCP-Ph] • "When you work with a patient that has got that influence within a community, particularly within an online community, then you can really get that product or that solution out there and you can build that trust because patients trust other patients like themselves." [P6-PE]
Novel approaches	
Patient experts (n=13, 65%)	• "I think some of the facilitators are now called patient entrepreneurs. We see this a lot in the area of diabetes because they are the experts. And you have to see a GP or specialist very regularly, and then you will quickly realize that you're the expert and not the physician is the expert." [P12-DE] • "And you see this happening more and more often, where patients are being treated as consultants." [P8-TP-Ph]
Patient engagement agencies (n=4, 20%)	• "But now with rising communities and patients' agencies that are coming up there and so on, I think we're now at the point where we are only able to get the voice into the complete process." [P5-PE-DE] • "There are now agencies as well that would help us connect in their indications." [P9-Ph]
Embedding them in the development teams (n=12, 60%)	• "But then I think they should be part of the team or at least of the advisory board so you don't lose focus of that. It's also kind of a strategic decision." [P12-DE] • "You would need patients right in there on the innovation teams driving the agenda." [P13-HCP-TP]
Expectation management (n=7, 35%)	• "You need to be very clear in terms of managing their expectations as to the context of what technology can, cannot do, what the aim is and that we're not going to solve everything here." [P14-TP-Ph] • "And it also helps manage expectations at this stage, to be really transparent and make sure that we understand each other and why, technically, some things are just not feasible, even though we wanted to do them." [P9-Ph]
Moderation services (n=6, 30%)	• "Sometimes part of the investment needs to be in professional facilitators to manage activities and also neutral facilitators are actually going to facilitate a more fair and open engagement process, which will reduce the risk of bias because there is that kind of risk that you might lead patients or your co-creative partners down a certain route that you hope the process is going to go." [P6-PE] • "If you really want to do this systematically, you probably have to work with an external partner who are experts in doing this. And I think that from what I have seen, a lot of startups, they kind of miss that point." [P12-DE]

Embedding Patients in All Cycles of the Development Process

To complete the picture, we systematically discussed each stage of the human-centered approach to design and asked each participant to assess the maturity of patient engagement in each of those phases on a scale of 1 to 5, with 1 being the least mature and 5 being the most mature. In addition to the usual 5 design phases, we asked the participants to assess the maturity of patient engagement in life cycle management. We added this step because, given the constant development of these tools, it is crucial to consider life cycle management; otherwise, they become obsolete in a couple of years if the developers do not cope with technological changes and the new tools that come to the market every day.

The early and middle stages of development were assessed as less mature, while later stages, such as prototyping and delivering the solution, were assessed as more mature, meaning that patients tend to be more involved in these stages of development. Patient engagement maturity was assessed as the lowest in life cycle management after the solution was delivered. Table 4 shows the average and SD of patient engagement maturity at each stage of the HCD, as assessed by the participants.

The maturity assessment was well aligned with participants' views on phases of the design where patients are most or least engaged, and the most common comment was that patients are usually engaged "too late" in the process, at a stage where it is difficult to make any radical changes to the design. The key themes and subthemes, their frequencies, and some sample quotes about phases in which patients are least or most involved in eHealth development are summarized in Table 5 for clarity.

Table 4. Maturity assessment of patient engagement in the key phases in human-centered design as assessed by the participants.

Phase of the human-centered design	Maturity assessment (out of 5), mean (SD)
Specify context	2.5 (0.8)
Define user requirements	2.7 (1.2)
Produce design	2.3 (0.9)
Prototype	3.3 (1.0)
Deliver solution	3.7 (0.8)
Life cycle management	2.2 (0.8)

Table 5. Phases where patients are least or most involved in eHealth development as expressed by the participants (N=20).

Theme	Sample quotes
Most involved	
At the end (too late; n=14, 70%)	• "And in general, the patient engagement happens a little too late in the process." [P11-PE] • "I think it's at testing, right. That's really no doubt that's the only place, which means there is already a prototype. It doesn't mean it's already done, but it's hard to change a prototype, right, at that stage." [P16-HCP-Ph] • "Basically, when the industry needs the patient for dissemination of an app, of support service, then they'll give it to the patient community, but very often too late." [P17-PA] • "If I take my experience, they were mainly involved at the end. Too late. Way too late." [P19-PE-DE] • "I think the main barrier is actually creating the product first and then looking for feedback second. It needs to be reversed. It needs to be the other way around." [P6-PE]
At the very beginning and very end (n=6, 30%)	• "I think some people tend to ask at the beginning, 'What are your needs?' And you do 10, 15 interviews at the very beginning. And then you develop what you think it has to be developed. And then, you ask them at the very end." [P19-PE-DE] • "But it's very often more probably—in the need definition, you have also more engagement, so before the design of the solution and then for the testing." [P2-PE-DE]
At the beginning (n=2, 10%)	• "I would say at the beginning." [P10-DE] • "So generally, I think patients tend to be involved in the early phase. When a company is coming up with the design, they kind of have been told, 'Oh, you need to do patient engagement.' So, they'll hold a focus group, they'll put some post it notes on a wall." [P15-TP]
Least involved	
In the middle (n=6, 30%)	• "I think they are least involved in the middle phases." [P19-PE-DE] • "this is a massive gap in the middle, where patients aren't being involved." [P15-TP]

We further asked them to brainstorm ideas that may help overcome the gaps in the areas where patients are least involved in the development process to ensure that they are embedded as cocreators in all development cycles. Figure 3 shows the

themes that emerged as a response to these 2 questions and their respective subthemes, reflecting the frequency of each theme (frequencies reflect the number of participants who mentioned that specific theme).

Participants determined that awareness and education is a key factor that may help overcome the current gaps in patient engagement in eHealth development. This encompasses awareness around the value that patient engagement may bring (7/20, 35%), about available patient engagement opportunities (3/20, 15%), and also about educating patients to equip them with the required skills for efficient patient engagement (3/20, 15%), educating eHealth providers about compliance and the right processes to compliantly engage patients in the development (2/20, 10%), and raising awareness about patient diversity and the differences between different patient profiles and skills (1/20, 5%).

Furthermore, a mindset shift is needed to enable the change. Shifting to a more patient-driven care model would encourage more patient engagement (7/20, 35%), fighting the stigma surrounding being a patient to empower patients to speak up more (2/20, 10%), and an organizational culture change on the solution providers' side to help embrace patient engagement (2/20, 10%).

Clearer regulatory guidance may also encourage solution providers to engage patients without being too worried about potential regulatory issues (8/20, 40%), incentivizing patients in a compliant manner would encourage patients and providers to partner together in the development process (8/20, 40%), building trust and transparency among the concerned stakeholders could also facilitate the collaboration (6/20, 30%), and finally developing frameworks to measure the value of patient engagement would help build the business case and encourage developers to dedicate the necessary resources for proper patient engagement (2/20, 10%).

The key themes and subthemes, their frequencies, and some sample quotes on how to overcome the gaps in patient engagement in eHealth development are summarized in Table 6 for clarity.

Figure 3. Gaps in patient engagement in eHealth development and how to overcome them. PE: patient engagement.

Table 6. Participants' suggestions on how to overcome the gaps in patient engagement in eHealth development (N=20).

Theme	Sample quotes
Awareness and education	
About patient engagement value (n=7, 35%)	• "Making sure that people understand and have hands-on experience of what's the value that patient engagement brings." [P16-HCP-Ph] • "Probably being more aware for the companies about the real benefit for the solution in having patients involved might be something that, let's say, can influence decision makers and having them participating." [P10-DE]
About PE[a] opportunities (n=3, 15%)	• "I think from a patient side, patients have no idea that these opportunities exist. So, you're not reaching the right patients. They don't know." [P7-PE-TP] • "Maybe also it's relevant to create awareness on the developer side how patients can be involved and over which platforms." [P20-PE-DE]
About PE skills (n=2, 10%)	• "I would argue training to make everybody far more adept at utilizing health research terminology, patient advocacy terms, and sort of the training capacity stuff that EUPATI is doing." [P11-PE] • "I think just getting that insight on what it is and how you do it, those are things which are trainable, right." [P14-TP-PH]
About compliance (n=2, 10%)	• "Just having a better understanding of the regulatory landscape. I'm not even asking to change the full regulatory landscape. I'm just asking for making it simple enough to be understood, so small companies that don't have these regulatory teams can actually understand it, and it's actionable, and they know what to do. And doing a lot more education." [P19-PE-DE] • "I've been helping the recruitment of global head of regulatory affairs recently; my role was to assess the patient engagement capacity. And the couple of people I interviewed, none of them were aware. They were delighted to see what I was sharing with them, but they were not aware. They didn't know what the EMA is doing." [P17-PA]
About patients' diversity (n=1, 5%)	• "I think it's actually promoting the perception, the difference between a lay and a patient expert." [P11-PE]
Mindset shift	
Shifting to a more patient-driven care (n=7, 35%)	• "I have a personal belief that actually the patient community should actually lead these efforts by putting out what are the unmet needs. And not as a response to briefing from one organization with one objective which is already pretty well-defined." [P17-PA] • "But just look at how that consumer journey happens in other industries and perhaps bring lessons learned from those other industries into health care. I think the consumer aspect is really key to look at other ways around how a patient will be engaging and expecting to have their journey improved." [P4-TP-PH]
Fighting the stigma (n=2, 10%)	• "This was really a time where you had to hide everything of being a chronic patient—people living with a chronic disease… I think it will, hopefully, get less stigma. And patients will be more and more able to speak up." [P5-PE-DE]
Organizational culture (n=2, 10%)	• "It takes a lot of courage for the organization to say, let's try out whatever the new tech players offer to not have to build the wheel all of the time and have a fast time to patient time to release on top of that." [P4-TP-Ph]
Clearer regulatory guidance (n=8, 40%)	• "Having more clarity around that. And it's almost like having clear guidelines within the health system, that this is- the way with clinical trials and so on." [P13-HCP-TP] • "To have some sort of, I don't even want to call it organization. It can even be a website. But somewhere where the regulatory landscape is easy and that they help you. Because that would help reduce the resistance and the fears inside of the companies." [P19-PE-DE]
Incentivizing patients (n=8, 40%)	• "Create the right level of incentive or engagement, where it's no longer a volunteer activity, but it is—let's just use the term, it's a clear contract with clear ins and outs." [P14-TP-Ph] • "I think it's important to incentivize people in some way because most people can't be bothered to give their feedback or—yeah, give a real motivation as to why people should get involved." [P7-PE-TP]

Theme	Sample quotes
Build trust and transparency (n=6, 30%)	• "It's this trust factor. So, building trust with patients, spending time with them, to align on how you want to work together." [P6-PE] • "Because participation is also built on trust and so everything has to be very transparent and clear." [P10-DE]
Frameworks to measure value (n=2, 10%)	• "We need to admit that whether you are for or against patient engagement until we measure the value and so on, it's kind of anti-discussion. Now it's time to actually measure what works, what doesn't work, measure the quality." [P17-PA] • "Could somebody quantify whether that actually has a statistical difference in the outcome? That's very, very hard to do because every tool is so different. But if you could quantify that, then you create a business case for leaders to invest in this direction." [P8-TP-Ph]

[a]PE: patient engagement.

Discussion

Principal Findings

This study shows that genuinely engaging patients in all phases of eHealth development can provide different types of value. The most prominent added value is that engaging patients since the early stages of development would help identify unmet needs, which is crucial because previous research showed that patient needs impact adoption, meaning that a tool that addresses a real need would be more successfully adopted [9,34]. Better usability and desirability are also outcomes of efficient patient engagement in addition to a better fit into the overall patient journey, another central factor for eHealth success, as highlighted by other researchers [35].

Ultimately, patient engagement leads to better eHealth adoption and stickiness, a fact pinpointed by 2 extensive systematic reviews that concluded that involving users in the development leads to better adoption [35,36]. Sustained adoption and tool stickiness may eventually lead to better health outcomes, as pointed out by other studies that highlighted the positive link between patient activation and treatment compliance [37] and showed an enhancement in patients' health outcomes and better quality of care with sustainable eHealth use [38,39]. In addition, considering that the digital divide is mostly considered a barrier to adoption [35,40], the inclusion of diverse profiles of patients with different skillsets and capabilities may contribute to the creation of more inclusive designs that lead to better equity and access to health care.

However, despite the invaluable contributions that patients may bring, the study participants agreed that they are usually engaged too late in the development process, mostly assuming a sounding role to test a ready-made prototype, as opposed to being embedded as equal partners and cocreators throughout the different phases of development. This aligns with the findings of a previous systematic review of international experiences that also determined that patient involvement is mostly achieved through consultation and that direct participation is less common [41].

This low engagement may be explained by some prominent barriers, notably compliance. The participants highlighted regulatory processes' complexity and lack of clarity as critical obstacles that hinder patient engagement. It was also noted that many stakeholders' perception of exaggerated regulatory risks creates resistance and reluctance to efficiently engage patients in the development. The challenge of compliance aligns with the findings of other researchers who pointed out negative attitudes toward engaging patients, especially concerning patient safety, as one of the key barriers to patient involvement [7,10].

The study participants also cited patient-related barriers, most prominently challenges in patient identification and access, primarily due to regulatory processes. In addition, factors such as health constraints depending on the patient's condition, as pointed out by previous research showing that a patient's health condition may impact how they engage with health care technologies [42] and sometimes lack the necessary skills for efficient engagement, were also recognized as potential hurdles.

Moreover, the power dynamics in the health care sector are generally not in patients' advantage, resulting in a general perception that they are more of passive receivers of care than equal partners in their health management. The economic model puts the power in the payers' hands, which are typically not the patients themselves but rather insurance companies; similarly, the decision power mostly lies in the clinicians' hands rather than the patients. This imbalance of power, paired with potential conflicts of interest among the key stakeholders, disfavors patients and results in situations where they find themselves not given a safe space to actively and equally contribute to discussions impacting their own health.

Other barriers include engaging patients in a "check-in-the-box" activity. Undertaken to look good on paper, without real essence, which is typically due to the lack of understanding of the value that genuine patient engagement may bring, as mentioned by other researchers, stressing that the lack of standardized best practices and metrics has made it challenging to achieve consistency and measure success in patient engagement [43]. Furthermore, embedding patients in the development process is a resource-intensive undertaking that requires time and budget that may not always be available. Besides mistrust issues, mainly driven by health data management concerns around eHealth tools, and the lack of transparency of some stakeholders, making it harder to gain patients' trust, an issue that has also been emphasized by other researchers [44,45].

On the positive side, the participants reflected on some facilitators that may enable better patient engagement. For

instance, working with engagement partners such as patient organizations may help overcome some of the regulatory hurdles. Clinicians can also play an active role in engagement, as noted by other researchers [46]. Engaging through the care team can help developers access not only patients who are already actively participating in patient organizations and advocacy groups but also most patients who are least active but may bring an essential perspective that would otherwise be missing. Active involvement in patient communities, offline and web-based, may also facilitate the collaboration between patients and their caregivers. Novel approaches, such as the rise of professional patient experts and patient engagement agencies, are also furthering the collaboration between the different stakeholders by simplifying the matchmaking process and helping overcome some patient identification and access barriers.

Practices that facilitate patient engagement include embedding the patients in the development teams, meaning hiring them as project managers or user experience experts if they have the necessary skill set. Suppose the involved patients do not necessarily have professional expertise and know-how to understand technical discussions. In that case, hiring professional moderation services that can help translate the language between patients and the development team and active expectation management, explaining what is possible or not possible, and what is in scope or not, can play a significant role in enabling successful collaboration between all parties.

Blueprint for Patient Engagement as Cocreators of eHealth Technologies

On the basis of the understanding of the value of patient engagement in eHealth development, its current state of maturity, and potential barriers and facilitators, we propose an end-to-end practical blueprint that can guide the relevant stakeholders to successfully engage patients as equal partners and cocreators in all phases of the HCD rather than mere testers of preplanned prototypes.

The first layer of the blueprint addresses sample considerations and the specific patient profiles that may best suit each phase of the HCD. Bearing in mind that the middle phases of the design are the least mature from a patient engagement perspective, partly because of the technical skills required to contribute to these phases efficiently, we suggest engaging with professional patient experts in the 3 middle phases. Working with patient experts ensures that the involved parties are equipped with the disease experience and the necessary know-how to engage in meaningful development discussions. The newly rising patient engagement agencies may help overcome patient identification and access barriers by matching the development teams with suitable patient experts. Organizations such as the European Patients' Academy on Therapeutic Innovation (EUPATI) also help match patient experts with health care researchers through their EUPATIConnect services [47].

However, it is crucial to warrant the diversity of the patients, especially in the first and last phases of the design, and to avoid working solely with patient experts in all phases. Involving lay patients in the mix and people who are not necessarily technically savvy will help the development team to create an inclusive design that is still usable even for the least capable users, enabling more health care equity and reducing the unbalancing effect of the digital divide. There are clinics and hospitals that have started establishing innovation laboratories, such as the University Hospital in Basel Switzerland [48], enabling the testing and cocreation of new health technologies with clinicians and lay patients who volunteered to test these tools.

The middle layer of the blueprint presents recommendations for the most suitable engagement approach for each phase. During the first phase of the design, the development team focused on specifying the context, including evidence review. Potential engagement approaches during this early phase may vary from monitoring discussions in web-based patient communities, looking directly into patient complaints or requests in clinics and hospitals, and conducting patient workshops or focus groups. One-to-one interviews may also provide in-depth insights, primarily when conducted in the patients' natural environment, to best reflect the entire patient journey and unaddressed needs.

As soon as the second phase of the design begins and the development team starts defining user requirements, the discussion becomes more technical. This is when professional moderation services of workshops and focus groups gain importance, as they can enhance the chances of a mutual understanding between stakeholders with varying technical skills. Ideation sessions using a design-thinking approach and benchmarking of existing apps are crucial tools at this stage.

The development teams sometimes merge the third and fourth phases of the design, producing the design and prototyping. However, it is worth noting that it may be worth starting with some A/B testing when producing the design, a way to compare 2 versions of a single variable, typically by testing a subject's response to variant A against variant B and determining which of the 2 variants is more effective [49]. For example, this approach may be used to test language and design elements before moving to prototyping. This saves time and effort by ensuring that the basic design resonates with the patients before producing the prototype. Simulations and laboratory and in-field testing are very relevant at this stage, as they help developers better understand actual user behavior rather than solely relying on self-reported feedback through surveys or checklists.

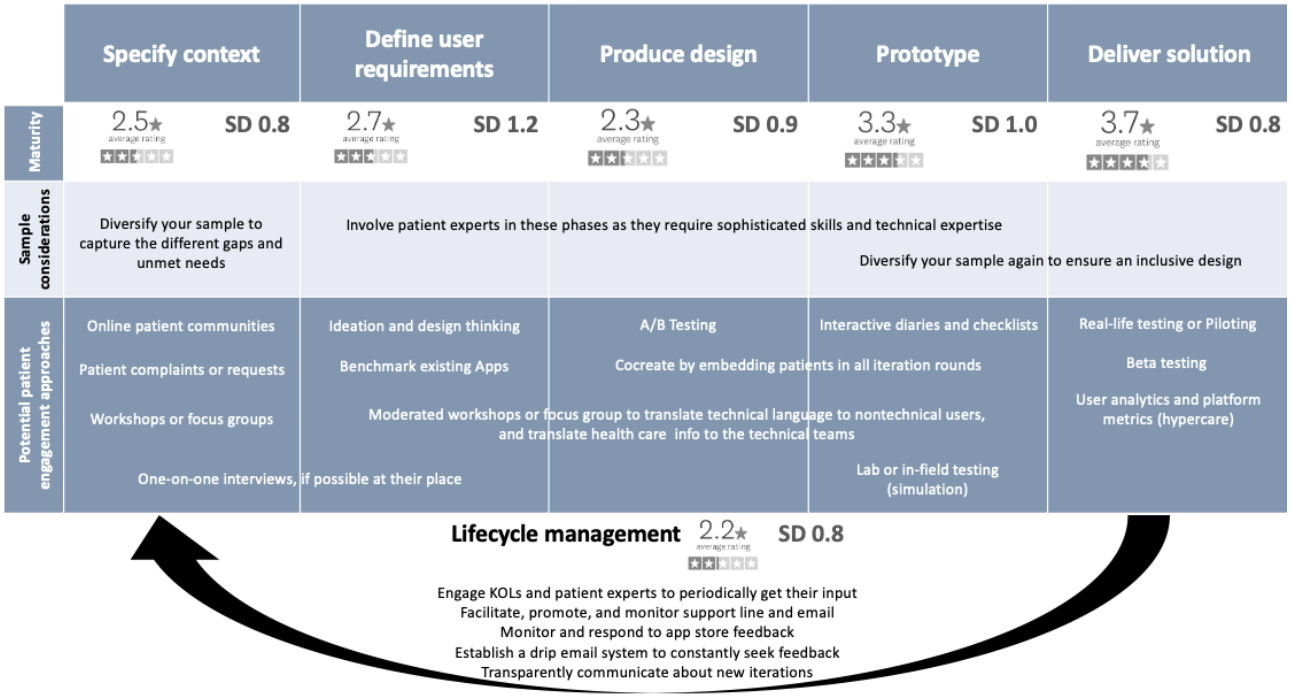
When delivering the tool, it is vital to test it in a real-life setting to ensure its fit into the patients' journey, meaning that it fits well into their daily routines and wholistic treatment plans. Beta-testing or piloting can be valuable in allowing developers to test their tool in a real-life setting on a smaller scale before rolling it out. It is also advised to have a hypercare period immediately after the launch of any eHealth tool, where developers closely monitor user analytics and platform metrics to act swiftly in case of any issues, providing a smoother integration in a real-world setting.

The bottom layer of the blueprint addresses the often-neglected life cycle management, which was assessed as the least mature from a patient engagement perspective. Ensuring an iterative approach that actively manages the constant development of

eHealth tools is critical for sustainable success, especially in an ever-changing technical landscape. Continually engaging with patients through consistent life cycle management ensures the stickiness and relevance of the tool. Engagement approaches can be as simple as actively monitoring and responding to the support line and email, or app-store feedback, but can also be more proactive, such as periodically engaging patient key opinion leaders to obtain their input. Other useful tools in this stage are drip email systems to constantly seek users' feedback and transparent communications about new iterations to inform users how their feedback was taken into account in the constant development of the tool.

Figure 4 shows the proposed blueprint for patient engagement in every phase of the HCD, presenting suggestions for patient sample considerations and recommendations for the most suitable engagement approaches for each phase.

Figure 4. Proposed blueprint for patient engagement as cocreators of eHealth technologies. KoL: key opinion leaders.



Practical Implications

Overcoming the current gaps in patient engagement in eHealth development requires consolidated efforts from all stakeholders in the complex health care ecosystem. Policy makers, clinicians, eHealth providers, pharmaceutical companies, insurance companies, patient organizations, advocacy groups, and health care innovation incubators must work hand in hand to induce change and harness the potential value that true cocreation with patients can bring.

Education and awareness are key to improving patient engagement. On the one hand, it involves educating patients and equipping them with the necessary knowledge and skills for effective engagement and contribution. Organizations such as EUPATI are already actively providing patient education programs [50]; however, more efforts are needed in the area of eHealth and all that it entails from specific technical skills. On the other hand, it is crucial to raise awareness of the value that patient engagement can bring, provide platforms that may help promote patient engagement opportunities, and provide more information about relevant compliance processes.

The study participants emphasized that there is a need for measurement frameworks that can help quantify the impact of patient engagement, as similarly highlighted in the systematic review by Bombard et al [51] and stressed by other scholars [52]. Some initiatives are digging deeper into this issue in attempts to create tools that may help evaluate patient engagement; for example, the Public and Patient Engagement Evaluation Tool developed at McMaster University in Canada [53,54] and its Norwegian expansion Evalueringsverktøy for Brukermedvirkning [55]. Furthermore, citizen-led organizations such as Patient-Focused Medicines Development and PARADIGM, a public-private partnership coled by the European Patients' Forum and The European Federation of Pharmaceutical Industries and Associations, are working on metrics that aim to help better monitor and evaluate patient engagement [56,57]. These efforts can help shed light on the business case for patient engagement to overcome the value perception barrier and enable genuine patient engagement as equal partners and cocreators.

Regulatory clarity and simplification would play a central role in facilitating patient engagement, given that compliance was deemed the most prominent barrier by study participants. A clear step-by-step approach, templates for agreements and contracts, clarity on patient remuneration, and health data privacy and management would encourage the relevant stakeholders to engage patients compliantly without being too worried about compliance risks. Owing to this lack of globally accepted guiding principles around patient involvement that identify and integrate good practices, organizations such as Patient-Focused Medicines Development work with patients and other stakeholders to cocreate frameworks and toolboxes that may guide the relevant stakeholders in their patient

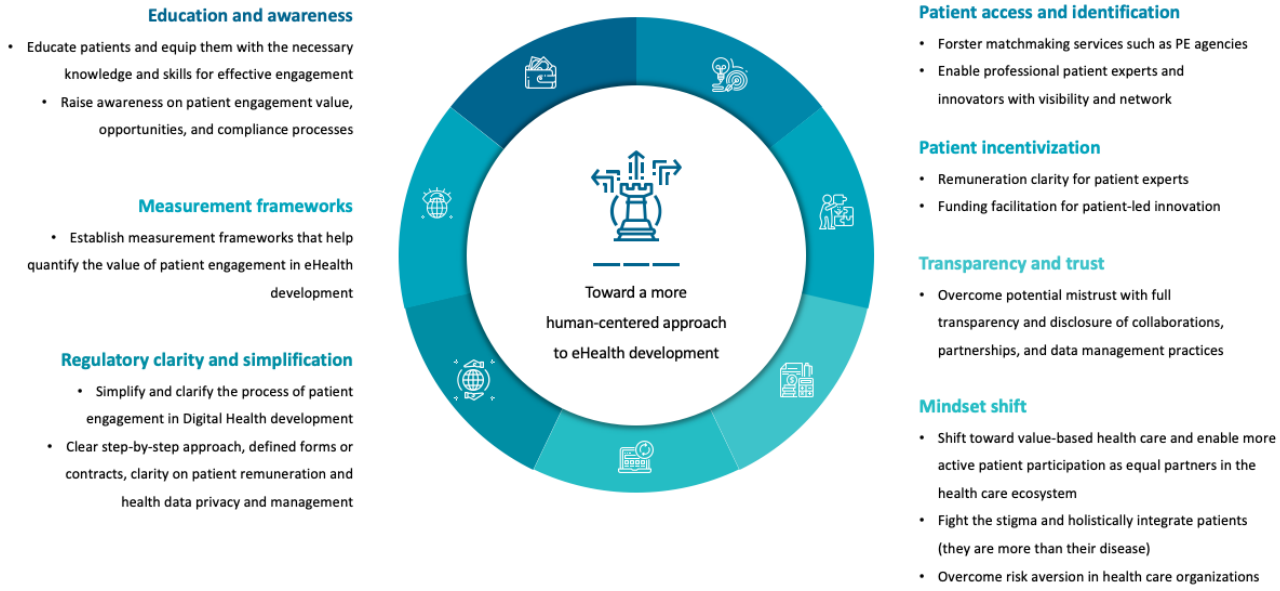
engagement efforts. Their patient engagement synapse provides sample agreements and contracts to facilitate compliant engagement and collaboration between patients and providers of medical technologies [58]. Another example of the type of guidance required is the Workgroup of European Cancer Patient Advocacy Networks Guiding Principles on Reasonable Legal Agreements between patient advocates and pharmaceutical companies [59]. This could also help solve the patient remuneration and incentivization dilemma as it offers clear guidance on the type of contract that may enable the collaboration in a compliant manner.

Barriers such as patient access and identification could be overcome by working with engagement partners and promoting novel approaches to empower professional patient experts and innovators with visibility and networks. The rise of matchmaking services can also play a positive role in overcoming these barriers; patient engagement agencies that focus on identifying and engaging with relevant patient experts to match them with suitable patient engagement opportunities are a good example. There are also organizations such as EUPATI that offer matchmaking services through their EUPATIConnect, as mentioned earlier [47]. Equally, we can overcome potential mistrust between different stakeholders with complete transparency and disclosure of collaborations and partnerships, as well as more clarity on data management practices.

This change would also require a mindset shift on several levels. First, a shift toward more value-based health care would help overcome the current imbalance in power dynamics and reinforce a more active role for patients; research has shown that only once patients are allowed to participate in managing their health actively, they take ownership of their disease management, thus improving health outcomes [60]. Second, it is key to fight the stigma around disease and being a patient, to empower patients, and to encourage them to speak up. Third, overcoming risk aversion toward patient engagement in health care organizations is a significant factor closely linked to regulatory clarity and simplification, as well as awareness of the value that genuine patient engagement may bring.

Figure 5 shows the key practical implications of this study and our recommendations for more patient-driven eHealth development.

Figure 5. Recommendations for more patient-driven eHealth development. PE: patient engagement.



Limitations and Recommendations for Future Research

This qualitative study has some limitations that we would like to outline. Our study was limited to 6 countries in a specific timeframe, and generalization to other settings that might have different characteristics, such as a different regulatory landscape, may be challenging. Moreover, the relatively small sample size and dynamic nature of eHealth necessitate a constant update of the findings to cope with the changes. Future research may address some of the cited limitations by covering other countries, timeframes, regulatory frameworks, and settings.

Conclusions

The outcome of this study contributes to creating awareness about the value of genuine patient engagement, barriers, and facilitators that impact engagement efforts and how to overcome the current gaps. We propose a blueprint that considers these specific findings and aims to facilitate the successful engagement of patients as cocreators in all phases of HCD rather than mere testers of preplanned prototypes.

Our findings highlight the tremendous value created by patient engagement in eHealth development. However, it also emphasizes the dominant gaps in the current patient involvement approaches. We shed light on the compelling obstacles behind these gaps and discuss ways to overcome them. It is important to note that overcoming the current gaps in patient engagement in eHealth development requires consolidated efforts from all stakeholders in the complex health care ecosystem, not only relying on medical technology providers to overcome them, as some factors go beyond their direct control.

Acknowledgments

This study was sponsored by Innosuisse (Swiss Innovation Agency, grant 60330.1 INNO-ICT). The role of Innosuisse is to promote science-based innovation in the interests of industry and society in Switzerland.

The research partner was the University of Applied Sciences, Northwestern Switzerland, in collaboration with the practice partner PersonalPulse GmbH, whose mission is to empower transformation in citizen-led health care innovation, working together with relevant stakeholders in the health care ecosystem to cocreate health care tools that are relevant, usable, and sustainable.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 137 KB - humanfactors_v9i4e41481_app1.pdf]

Multimedia Appendix 2
Survey Monkey form.
[PDF File (Adobe PDF File), 203 KB - humanfactors_v9i4e41481_app2.pdf]

Multimedia Appendix 3
Phases of thematic analysis after Braun and Clarke [28,29].
[PDF File (Adobe PDF File), 48 KB - humanfactors_v9i4e41481_app3.pdf]

Multimedia Appendix 4
Barriers to patient engagement as expressed by the participants.
[PDF File (Adobe PDF File), 118 KB - humanfactors_v9i4e41481_app4.pdf]

References

1. Wilson J. Breaking down barriers to patient engagement. Br J Nurs 2010;19(8):473. [doi: 10.12968/bjon.2010.19.8.47663] [Medline: 20505610]
2. Manafo E, Petermann L, Mason-Lai P, Vandall-Walker V. Patient engagement in Canada: a scoping review of the 'how' and 'what' of patient engagement in health research. Health Res Policy Syst 2018 Feb 07;16(1):5 [FREE Full text] [doi: 10.1186/s12961-018-0282-4] [Medline: 29415734]
3. Kovacs Burns K, Bellows M, Eigenseher C, Gallivan J. 'Practical' resources to support patient and family engagement in healthcare decisions: a scoping review. BMC Health Serv Res 2014 Apr 15;14:175 [FREE Full text] [doi: 10.1186/1472-6963-14-175] [Medline: 24735787]
4. Gallivan J, Kovacs Burns KA, Bellows M, Eigenseher C. The many faces of patient engagement. J Particip Med 2012 Dec 26;4(23):e32.
5. Higgins T, Larson E, Schnall R. Unraveling the meaning of patient engagement: a concept analysis. Patient Educ Couns 2017 Jan;100(1):30-36. [doi: 10.1016/j.pec.2016.09.002] [Medline: 27665500]
6. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: a systematic review. BMC Health Serv Res 2014 Feb 26;14:89 [FREE Full text] [doi: 10.1186/1472-6963-14-89] [Medline: 24568690]
7. Gagliardi AR, Lemieux-Charles L, Brown AD, Sullivan T, Goel V. Barriers to patient involvement in health service planning and evaluation: an exploratory study. Patient Educ Couns 2008 Feb;70(2):234-241. [doi: 10.1016/j.pec.2007.09.009] [Medline: 18023129]
8. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet 2014 Jan 11;383(9912):156-165. [doi: 10.1016/S0140-6736(13)62229-1] [Medline: 24411644]
9. Aguayo GA, Goetzinger C, Scibilia R, Fischer A, Seuring T, Tran VT, et al. Methods to generate innovative research ideas and improve patient and public involvement in modern epidemiological research: review, patient viewpoint, and guidelines for implementation of a digital cohort study. J Med Internet Res 2021 Dec 23;23(12):e25743 [FREE Full text] [doi: 10.2196/25743] [Medline: 34941554]
10. Chegini Z, Janati A, Babaie J, Pouraghaei M. Exploring the barriers to patient engagement in the delivery of safe care in Iranian hospitals: a qualitative study. Nurs Open 2020 Jan;7(1):457-465 [FREE Full text] [doi: 10.1002/nop2.411] [Medline: 31871731]

11. Geissler J, Ryll B, di Priolo SL, Uhlenhopp M. Improving patient involvement in medicines research and development:: a practical roadmap. Ther Innov Regul Sci 2017 Sep;51(5):612-619. [doi: 10.1177/2168479017706405] [Medline: 30231692]
12. Chan K. A design thinking mindset beyond the public health model. World Med Health Policy 2018 Mar;10(1):111-119. [doi: 10.1002/wmh3.253]
13. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
14. Van der Bijl-Brouwe M. The challenges of human-centred design in a public sector innovation Contex. In: Future Focused Thinking - DRS International Conference 2016. 2016 Presented at: DRS '16; June 27-30, 2016; Brighton, UK. [doi: 10.21606/drs.2016.294]
15. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
16. Hooper D. User-centered design for medical devices: if it isn't documented it doesn't exist. Medical Device and Diagnostic Industry. 2012 Sep 7. URL: https://www.mddionline.com/design-engineering/user-centered-design-medical-devices-if-it-isnt-documented-it-doesnt-exist [accessed 2022-10-04]
17. Backes C, Moyano C, Rimaud C, Bienvenu C, Schneider MP. Digital medication adherence support: could healthcare providers recommend mobile health apps? Front Med Technol 2021 Feb 17;2:616242 [FREE Full text] [doi: 10.3389/fmedt.2020.616242] [Medline: 35047896]
18. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. Boca Raton, FL, USA: Lawrence Erlbaum Associates; 1986.
19. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. International Organization for Standardization Technical Committee. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2022-10-04]
20. Meyer A, Rose DH. Universal design for individual differences. Educ Leadership 2000 Nov;58(3):39-43 [FREE Full text]
21. The Field Guide to Human-Centered Design. Design Kit. 2015. URL: https://www.designkit.org//resources/1 [accessed 2022-10-04]
22. What is Design Thinking? Hasso Plattner Institute (HPI). 2022. URL: https://hpi.de/school-of-design-thinking/design-thinking/was-ist-design-thinking.html [accessed 2022-10-04]
23. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. BMJ 2000 Jan 01;320(7226):50-52 [FREE Full text] [doi: 10.1136/bmj.320.7226.50] [Medline: 10617534]
24. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995 Jul 01;311(6996):42-45 [FREE Full text] [doi: 10.1136/bmj.311.6996.42] [Medline: 7613329]
25. eHealth. World Health Organization. 2022. URL: http://www.emro.who.int/health-topics/ehealth/ [accessed 2022-10-04]
26. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA, USA: Sage Publications; 2013.
27. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks, CA, USA: Sage Publications; 2017.
28. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014 Oct 16;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Yin RK. Case Study Research and Applications: Design and Methods. 6th edition. Thousand Oaks, CA, USA: Sage Publications; 2017.
31. Bryman A, Bell E. Business Research Methods. 4th edition. Oxford, UK: Oxford University Press; 2017.
32. PersonalPulse. 2022. URL: https://personalpulse.com/ [accessed 2022-10-04]
33. Knox S, Burkard AW. Qualitative research interviews. Psychother Res 2009 Jul;19(4-5):566-575. [doi: 10.1080/10503300802702105] [Medline: 19579087]
34. Connolly SL, Sullivan JL, Lindsay JA, Shimada SL, Heyworth L, Weaver KR, et al. Factors influencing uptake of telemental health via videoconferencing at high and low adoption sites within the Department of Veterans Affairs during COVID-19: a qualitative study. Implement Sci Commun 2022 Jun 20;3(1):66 [FREE Full text] [doi: 10.1186/s43058-022-00318-x] [Medline: 35725642]
35. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 05;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
36. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]

37. Zhang X, Zhang R, Lu X. Exploring the effects of patient activation in online health communities on patient compliance. Telemed J E Health 2020 Nov;26(11):1373-1382. [doi: 10.1089/tmj.2019.0258] [Medline: 32017672]
38. Zocchi MS, Robinson SA, Ash AS, Vimalananda VG, Wolfe HL, Hogan TP, et al. Patient portal engagement and diabetes management among new portal users in the Veterans Health Administration. J Am Med Inform Assoc 2021 Sep 18;28(10):2176-2183 [FREE Full text] [doi: 10.1093/jamia/ocab115] [Medline: 34339500]
39. Schmalz O, Jacob C, Ammann J, Liss B, Iivanainen S, Kammermann M, et al. Digital monitoring and management of patients with advanced or metastatic non-small cell lung cancer treated with cancer immunotherapy and its impact on quality of clinical care: interview and survey study among health care professionals and patients. J Med Internet Res 2020 Dec 21;22(12):e18655 [FREE Full text] [doi: 10.2196/18655] [Medline: 33346738]
40. Connolly SL, Stolzmann KL, Heyworth L, Sullivan JL, Shimada SL, Weaver KR, et al. Patient and provider predictors of telemental health use prior to and during the COVID-19 pandemic within the Department of Veterans Affairs. Am Psychol 2022;77(2):249-261. [doi: 10.1037/amp0000895] [Medline: 34941310]
41. Gagnon MP, Tantchou Dipankui M, Poder TG, Payne-Gagnon J, Mbemba G, Beretta V. Patient and public involvement in health technology assessment: update of a systematic review of international experiences. Int J Technol Assess Health Care 2021 Feb 05;37:e36. [doi: 10.1017/S0266462321000064] [Medline: 33541449]
42. Sezgin E, Weiler M, Weiler A, Lin S. Proposing an ecosystem of digital health solutions for teens with chronic conditions transitioning to self-management and independence: exploratory qualitative study. J Med Internet Res 2018 Sep 06;20(9):e10285 [FREE Full text] [doi: 10.2196/10285] [Medline: 30190253]
43. Pushparajah DS. Making patient engagement a reality. Patient 2018 Feb;11(1):1-8 [FREE Full text] [doi: 10.1007/s40271-017-0264-6] [Medline: 28741235]
44. Cosmi F, Mariottoni B, Tarquini B, D'Orazio S, Bennati M, Morganti M, et al. [The patient in clinical research: disposable guinea pig or involved actor?]. G Ital Cardiol (Rome) 2020 Apr;21(4):309-314. [doi: 10.1714/3328.32993] [Medline: 32202565]
45. Powell RE, Doty A, Casten RJ, Rovner BW, Rising KL. A qualitative analysis of interprofessional healthcare team members' perceptions of patient barriers to healthcare engagement. BMC Health Serv Res 2016 Sep 20;16:493 [FREE Full text] [doi: 10.1186/s12913-016-1751-5] [Medline: 27644704]
46. Bright FA, Kayes NM, Worrall L, McPherson KM. A conceptual review of engagement in healthcare and rehabilitation. Disabil Rehabil 2015;37(8):643-654. [doi: 10.3109/09638288.2014.933899] [Medline: 24969698]
47. EUPATIConnect. 2022. URL: https://connect.eupati.eu/ [accessed 2022-10-04]
48. Innovation Lab. University Hospital Basel Switzerland. 2022. URL: https://innovation.usb.ch/index/innovation-lab.html [accessed 2022-10-04]
49. Kohavi R, Tang D, Xu Y. Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge, UK: Cambridge University Press; 2020.
50. EUPATI Training Portfolio. EUPATI. 2022. URL: https://eupati.eu/training/ [accessed 2022-10-04]
51. Bombard Y, Baker GR, Orlando E, Fancott C, Bhatia P, Casalino S, et al. Engaging patients to improve quality of care: a systematic review. Implement Sci 2018 Jul 26;13(1):98 [FREE Full text] [doi: 10.1186/s13012-018-0784-z] [Medline: 30045735]
52. Hoos A, Anderson J, Boutin M, Dewulf L, Geissler J, Johnston G, et al. Partnering with patients in the development and lifecycle of medicines: a call for action. Ther Innov Regul Sci 2015 Nov;49(6):929-939 [FREE Full text] [doi: 10.1177/2168479015580384] [Medline: 26539338]
53. Abelson J, Tripp L, Kandasamy S, Burrows K, PPEET Implementation Study Team. Supporting the evaluation of public and patient engagement in health system organizations: results from an implementation research study. Health Expect 2019 Oct;22(5):1132-1143 [FREE Full text] [doi: 10.1111/hex.12949] [Medline: 31373754]
54. Abelson J, Li K, Wilson G, Shields K, Schneider C, Boesveld S. Supporting quality public and patient engagement in health system organizations: development and usability testing of the Public and Patient Engagement Evaluation Tool. Health Expect 2016 Aug;19(4):817-827 [FREE Full text] [doi: 10.1111/hex.12378] [Medline: 26113295]
55. Garratt A, Sagen J, Børøsund E, Varsi C, Kjeken I, Dagfinrud H, et al. The public and patient engagement evaluation tool: forward-backwards translation and cultural adaption to Norwegian. BMC Musculoskelet Disord 2022 Jun 09;23(1):556 [FREE Full text] [doi: 10.1186/s12891-022-05514-3] [Medline: 35676674]
56. Medicine and device lifecycles, digital health, data and health systems for better outcomes. Patient Focused Medicines Development. 2022. URL: https://patientfocusedmedicine.org/ [accessed 2022-10-04]
57. Patient Engagement Monitoring and Evaluation Framework. PARADIGM. 2022. URL: http://imi-paradigm.eu/petoolbox/monitoring-evaluation/ [accessed 2022-10-04]
58. Template Agreement between a patient or patient advocacy organization and a medical technology/device company. Patient Engagement Synapse, Patient Focused Medicines Development. 2021. URL: https://patientengagement.synapseconnect.org/resources/template-agreement-between-a-patient-or-patient-advocacy-organization-and-a-medical-technology-device-company-1 [accessed 2022-10-04]

59. Guiding Principles on Reasonable Agreements between Patient Advocates and Pharmaceutical Companies. Workgroup of European Cancer Patient Advocacy Networks (WECAN). 2018. URL: https://wecanadvocate.eu/wp-content/uploads/2019/03/Guiding-Principles_final-document6.2_clean.pdf [accessed 2022-10-04]
60. Soto C, Strain WD. Tackling clinical inertia: use of coproduction to improve patient engagement. J Diabetes 2018 Dec;10(12):942-947. [doi: 10.1111/1753-0407.12814] [Medline: 29931752]

Abbreviations

EUPATI: European Patients' Academy on Therapeutic Innovation
HCD: human-centered design

Edited by A Kushniruk; submitted 03.08.22; peer-reviewed by C Varsi; comments to author 24.08.22; revised version received 30.08.22; accepted 10.09.22; published 06.10.22.

Please cite as:
Jacob C, Bourke S, Heuss S
From Testers to Cocreators—the Value of and Approaches to Successful Patient Engagement in the Development of eHealth Solutions: Qualitative Expert Interview Study
JMIR Hum Factors 2022;9(4):e41481
URL: https://humanfactors.jmir.org/2022/4/e41481
doi:10.2196/41481
PMID:36102548

Original Paper

Motivational Interviewing Conversational Agent for Parents as Proxies for Their Children in Healthy Eating: Development and User Testing

Diva Smriti[1], BDES, MSc; Tsui-Sui Annie Kao[2], FNP-BC, RN, PhD; Rahil Rathod[3], BE, MSc; Ji Youn Shin[4], BFA, MDES, PhD; Wei Peng[5], BA, MA, PhD; Jake Williams[1], BA, MSc, PhD; Munif Ishad Mujib[1], BSc, PhD; Meghan Colosimo[6], BSc, MSc, PhD; Jina Huh-Yoo[1], BA, MHCI, PhD

[1]College of Computing and Informatics, Drexel University, Philadelphia, PA, United States

[2]College of Nursing, Michigan State University, East Lansing, MI, United States

[3]Tata Consultancy Services, Edison, NJ, United States

[4]College of Design, University of Minnesota, Minneapolis, MN, United States

[5]College of Communication Arts and Sciences, Michigan State University, East Lansing, MI, United States

[6]Vivante Health, Houston, TX, United States

Corresponding Author:
Diva Smriti, BDES, MSc
College of Computing and Informatics
Drexel University
3675 Market Street
10th floor
Philadelphia, PA, 19104
United States
Phone: 1 2158952474
Email: ds3659@drexel.edu

Abstract

Background: Increased adoption of off-the-shelf conversational agents (CAs) brings opportunities to integrate therapeutic interventions. Motivational Interviewing (MI) can then be integrated with CAs for cost-effective access to it. MI can be especially beneficial for parents who often have low motivation because of limited time and resources to eat healthy together with their children.

Objective: We developed a Motivational Interviewing Conversational Agent (MICA) to improve healthy eating in parents who serve as a proxy for health behavior change in their children. Proxy relationships involve a person serving as a catalyst for behavior change in another person. Parents, serving as proxies, can bring about behavior change in their children.

Methods: We conducted user test sessions of the MICA prototype to understand the perceived acceptability and usefulness of the MICA prototype by parents. A total of 24 parents of young children participated in 2 user test sessions with MICA, approximately 2 weeks apart. After parents' interaction with the MICA prototype in each user test session, we used qualitative interviews to understand parents' perceptions and suggestions for improvements in MICA.

Results: Findings showed participants' perceived usefulness of MICAs for helping them self-reflect and motivating them to adopt healthier eating habits together with their children. Participants further suggested various ways in which MICA can help them safely manage their children's eating behaviors and provide customized support for their proxy needs and goals.

Conclusions: We have discussed how the user experience of CAs can be improved to uniquely offer support to parents who serve as proxies in changing the behavior of their children. We have concluded with implications for a larger context of designing MI-based CAs for supporting proxy relationships for health behavior change.

(JMIR Hum Factors 2022;9(4):e38908) doi:10.2196/38908

KEYWORDS

conversational agents; voice user interface; voice agents; proxy; motivational interviewing; parents; healthy eating

Introduction

Motivation

Off-the-shelf conversational agents (CAs) such as Google Home, Amazon Alexa, and Echo are becoming increasingly popular for managing health in everyday settings. More than 50 million purchases of Amazon Echo have been made in the United States alone [1]. CAs have also been shown to help in improving parenting practices. CAs can help set up children's pre-established healthy diet–based plans as well as help track children's food consumption [2]. Despite the benefits of CAs in supporting a family's healthy eating habits, studies have also highlighted limitations that still need to be addressed. A study on the perception of mealtime technologies showed that parents did not like the idea of technology telling them or their children what to do and wanted more control over their family's food decisions [3]. Furthermore, parents believed the use of CAs would increase the dependence of their families on technology [3]. These findings point to the need for technologies such as CAs to support parental control and involvement in the meal decisions of their children.

Approximately 40% of US children's daily caloric intake consists of high-fat foods or sugary drinks, raising concerns about linked chronic illnesses due to unhealthy eating [4]. Parents perform a critical role in shaping children's eating habits and are often the decision makers for the family meal [5-7], which poses a long-term health effect as adults. Parents take on the responsibility and accountability of bringing about a health behavior change in their children that they care for. This form of relationship, which we call a *proxy relationship*, is a socially mediated relationship that involves giving a person—the proxy—who supports the target individual of behavior change the power to achieve the target individual's goals [8,9]. Studies on CAs for family use often focus on direct interactions with the target individual, treating behavior change as an individual problem. However, in the case of children who may not have access to or know how to correctly use CAs, parents can interact with CAs and can serve as a proxy for behavior change in their children.

For parents serving as proxies, factors such as lack of motivation, time, and work often get in the way of parents practicing healthy eating habits together with their children [10-12]. The Motivational Interviewing (MI) technique has emerged as an effective counseling method that helps individuals discover motivations and strategies for personalized behavior change [13], especially for diet modification [14-16]. MI emphasizes individual autonomy and helps individuals self-form personalized solutions that might work for their physical, social, and economic constraints [13,17]. Technology-based MI approaches deliver adaptations of MI using technology and various types of media including CAs [18]. These technology-based approaches have been shown to extend the MI intervention beyond what a therapist could offer face-to-face and provide cost-effective access to therapeutic services to underserved clients, such as rural populations [19]. However, current technology-based MIs are focused on direct interactions with the target individual of behavior change. This approach might not be as useful for those target individuals with low literacy, agency, or little motivation who may not or cannot interact with technology directly. This limitation of MI-based technologies is especially true for young children, who are often dependent on their parents to make decisions on their behalf. Behavior change in this context is a social problem; however, existing MI-based CAs often overlook this social context of behavior change. More work is required to design CAs for addressing behavior change problems that require more than one individual to work together (eg, parent and child), where one individual is dependent on the other to make decisions on their behalf, for instance.

In this study, we investigated how MI-based CAs can help address the gap in supporting health behavior change facilitated by proxy relationships. To explore the requirements for this design opportunity, we developed a working prototype of an MI-based CA called Motivational Interviewing Conversational Agent (MICA). The MICA prototype incorporates automated MI to help parents of young children adopt healthy eating practices. As parents are motivated to eat healthy using MICA, they will be better supported to inculcate similar healthy habits in their children. The end goal of MICA is to support parents serving as proxies for their children in healthy eating through cost-effective access to MI. Our objectives were to understand (1) how parents perceive MICA to best deliver its interventions to help them serve as proxies for health behavior change for their children and (2) how parents experience a CA, such as MICA, for behavior change to understand how families use in-home CAs for managing everyday practices. Our study is a starting point toward MI-based CAs supporting proxy relationships wherein the proxy individual facilitates a behavior change in the target individual.

Background

Parents as Proxies for Behavior Change in Their Children

A proxy relationship is a socially mediated relationship that involves a person (proxy) having the power to achieve another person's goals [8,9]. Proxies often help individuals rely on them to accomplish goals when the individuals (1) do not have the adequate means or resources to reach their goals, (2) have more trust in the proxy than themselves to lead them toward behavior change, and (3) do not want to be accountable for reaching their goals. This proxy relationship allows target individuals of behavior change to make better decisions, invest appropriate knowledge and time, and be accountable for managing their health. Target individuals of behavior change may also lack the ability to make informed choices during their behavior change process because of cognitive or physical disability or lower technical and health literacy. For instance, parents serving as a proxy and helping their children in health behavior change can be especially beneficial for the child where the child may not be able to make informed decisions. Parents also play a critical role in how children establish their long-term eating habits from a young age [7,20,21]. Poor diet during childhood leads to obesity, which is also highly linked with chronic illnesses, including diabetes and cardiovascular diseases [22].

Despite the critical role parents play, they often have the extra burden of not getting enough time and support for themselves [23]. Supporting parents can help reduce their burden and lead to extended periods of parenting. Researchers have identified challenges for parents in the form of a negative psychological response to the obligations of being a parent [24]. Such a response can negatively affect the motivation and self-esteem of parents to adopt healthy behaviors as a family [25]. A scoping review of health-related parenting showed that the areas where parents require support are accessing complex information, guiding through decision-making processes, or self-monitoring with appropriate feedback [26]. Parenting programs help parents in mitigating these challenges and motivate parents to provide better support to their children. However, parents face barriers to participating in parenting programs that are highly correlated with their socioeconomic status, material hardship, and resource limitations (eg, transportation issues and childcare needs) [27-29]. Such cost-related barriers can limit parents' access to and use of the available resources.

Emerging technologies, such as CAs, have the potential of providing support to parents in a cost-effective manner, suited to the needs of parents and their children [30-32]. However, in the case of parents wanting to eat healthier together with their children, direct interaction of the technology with the children may not be possible or practical because of the constraints around technology accessibility, literacy, or accountability. More work is required to understand how technology can work with parents to help them serve as a proxy and change the health behavior of their children.

In-home CA Use for Parenting

For the last few years, an increasing number of studies have looked at how in-home CAs can support a wide array of tasks for improving family practices. As off-the-shelf CAs became available on personal devices, people can adopt them at reasonable costs compared with the early forms of laboratory-based CAs. CAs can explain complex concepts using simple communication techniques without the technical literacy constraints that web-based patient portals generally require [33]. CAs also help users verbalize their goals for behavior change [34]. CAs offer the potential of customization to fit better the needs and goals of individuals with low health and technical literacy [30,35].

Several studies showed that in-home CAs can augment family practices by helping parents perform and regulate day-to-day tasks [2,36,37]. For instance, field studies by Beneteau et al [36,37] showed that in-home CAs facilitated collaborative learning experiences in families. Researchers have also conducted content analysis and collected use data of households using in-home CAs. The results showed how agents are integrated into the family's daily routines and used for varied purposes by different family members [38-41]. However, CAs lacked contextual knowledge of the households they were placed in and the activities that happened around them, often leading to privacy concerns [41]. In addition, CAs are not best equipped for children because parents have to be heavily involved, at least in the initial stage of CA use. For example, parents have to teach children how to effectively use commands and communication styles to use the CA better and safely [2,42,43].

A few recent studies examined how CAs can support healthy eating in parent-child dyads. For example, Garg and Sengupta [2] discussed that parents could use in-home CAs to track children's food consumption and nudge children to eat healthier according to their rules. Jo et al [44] conducted a field trial of a sensor-based speech recognition system named MAMAS for promoting healthy eating behaviors in parents and children. They found that the use of MAMAS helped promote autonomy in children's eating behaviors and positively affected family eating practices. These studies showed the benefits of family-based CAs in establishing household rules and values and the potential CAs have in improving children's learning and eating habits. However, Chen et al [3] found that parents were skeptical about using CAs at mealtime for their children, as they perceived it to be distracting and intrusive. Chen et al [3] highlighted the importance of designing CAs in partnership with parents to account for their needs and trust. These studies did not focus on motivating behavior change because the CAs currently available in the market do not support goal-specific or context-specific conversations. In addition, these studies focus on the direct interaction of the CA with the target audience for behavior change; that is, either the parents or the children. However, in the case where the target individuals of behavior change are constrained by technology accessibility, literacy, or accountability, such as children, direct interaction of the CA with the target individuals of behavior change may not be possible or practical. More work is required to understand how CAs can help with parent-child proxy relationships.

Healthy Eating Through MI in Parenting Context

Among many factors that influence children's eating patterns, parents' feeding styles [45] and modeling parents' preferences, intake, and acceptance of food play a significant role [46]. Accordingly, involving parents as core participants in interventions improves children's healthy eating habits [22]. However, because of the complexities with parenting [12], lack of time [10], and maternal stress [11], even with parents' desire to eat healthier as a family, motivation and confidence to eat healthier can be hindered [47].

MI, as a psychological therapeutic approach, provides a solution to this challenge by emphasizing individual motivation and autonomy. MI helps individuals self-form personalized solutions that might work for their physical, social, and economic constraints [13,17]. The MI technique has emerged as an effective counseling model for healthy eating and diet modification [15]. MI is a person-centered approach to help clients resolve barriers to motivation, reduce resistance, and foster commitment to lifestyle changes in modifying healthy behavior [16], especially when the change requires resolving various personal complexities [14]. In supporting effective long-term health behavior changes such as healthy eating, modified interventions such as MI are required that meet the needs of each individual's behavior change stage at the time [48]. The transtheoretical model (TTM) of behavior change is an integrative theory of therapy that assesses individuals' readiness and confidence to make changes to new healthier

behavior and provides strategies accordingly [49]. The TTM has been instrumental in developing the MI technique, as the intrinsic motivation required to move through the different behavior change stages of the TTM is gained from the MI technique [50]. The TTM has been applied to a wide array of health and wellness contexts [51] with the MI technique to guide behavior change intervention design in the context of eating habits, such as women's binge eating [52], eating disorders [53], and vegetable and fruit consumption [48].

Several studies investigated applying MI as part of the TTM in the context of parent-child dyads to examine the factors that influence parental readiness to facilitate their children's weight management, including diet and physical activities. For example, Rhee et al [54] showed that health care providers' suggestions based on MI or behavior change strategies facilitated parental readiness for children's dietary behaviors. Another study also aimed to identify characteristics associated with parental confidence to change their family's weight management behaviors, including eating and physical activities [55]. The results emphasized the importance of providing appropriate, targeted counseling for parents from the early stages, as overweight-related behaviors of children begin early in childhood. Similarly, in their focus group study, Bolling et al [56] discussed parents' perspectives on effective approaches to managing their children's weight that health care providers could make use of in their counseling. The results indicated that parents desired earlier, direct, and explicit interventions and suggestions. Studies suggest that earlier interventions involving early, direct counseling were positively associated with parents' readiness and confidence to change their children's dietary and weight management behaviors. In this regard, it becomes beneficial to introduce MI in the earlier stages of the parent-child dyad's eating behavior. MICA incorporates MI starting from the first session to account for an early intervention incorporating strategies that parents come up with themselves to align with their goals. MI thus provides a way for CAs to take into account the goals and needs of parents supporting their children in the conversations for improving the TTM. As MI brings numerous benefits for parents supporting their children in behavior change, it becomes essential to understand how MI can help parents serve as proxies for their children.

Although studies highlighted the usefulness of in-home CAs in the parenting context, they mainly investigated the effectiveness of functions implemented in commercial products. The design requirements of CAs that support particular health behaviors for parent-child dyads, wherein the parents serve as proxies for their children, were not actively discussed. As families use in-home CAs for managing everyday practices, it is imperative to understand how parents experience a CA for behavior change. Furthermore, it is vital to understand how parents perceive it can best deliver its interventions to help them serve as proxies for health behavior change for their children.

Methods

Overview

We developed a working prototype of a MICA and conducted a user test remotely in a laboratory setting using semistructured interviews. We conducted 44 user test sessions of the MICA prototype remotely in a laboratory setting for 2 weeks with 24 parents of young children interested in healthy eating with their families. We used semistructured interviews to understand parent participants' acceptance of using the MICA prototype and design opportunities to improve the MICA prototype. The goals of this study were to (1) understand parents' perceived usefulness of the system in helping them change their own eating behaviors that can influence their children's eating habits, (2) envision the ways in which MICA can help parents support their children to eat healthier, and (3) formatively evaluate how to improve the experience of using MICA to better support parents as proxies.

MI-Based Conversation Script of MICA

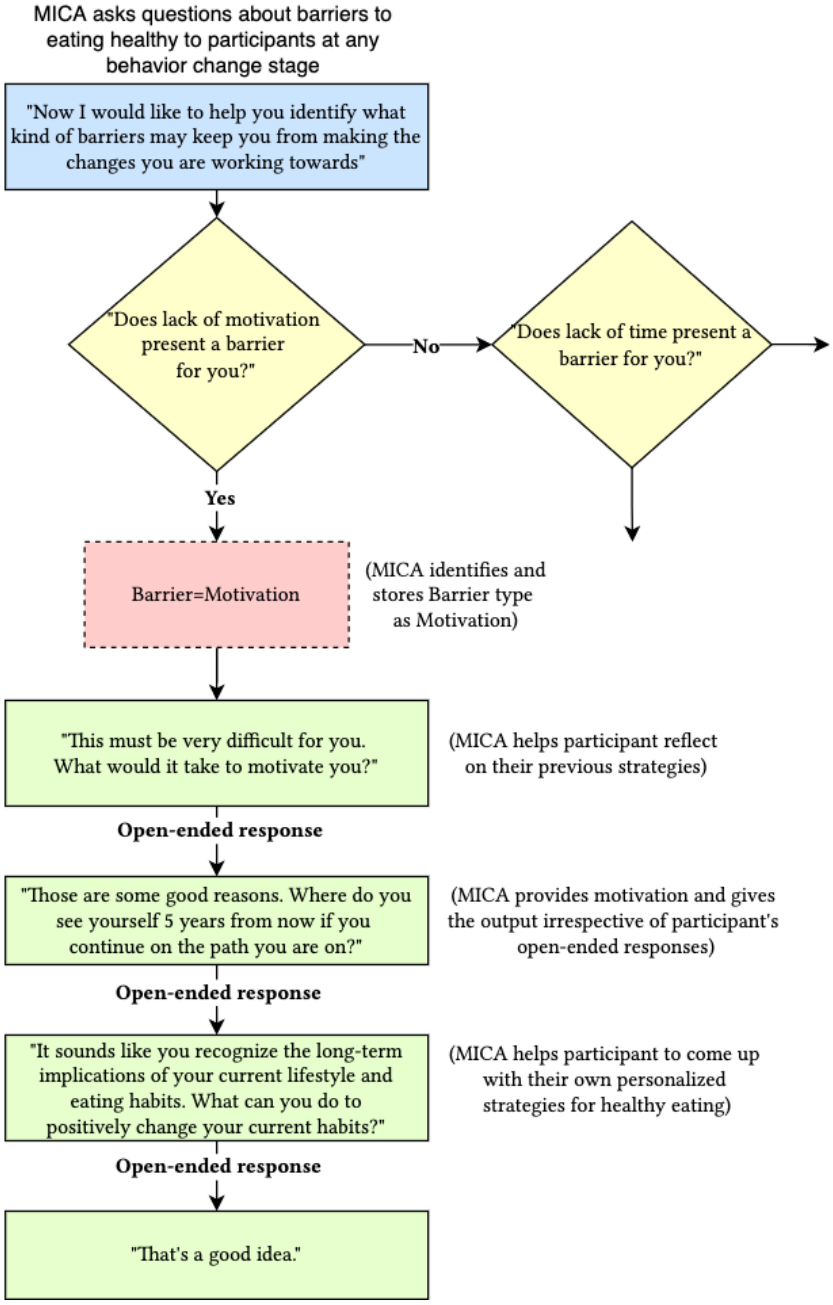
The current conversation script that MICA uses to converse with parents is based on MI scripts that human MI therapists follow. The MI script follows a logic tree structure (Figure 1). The MI script branches out into different subtopic questions based on the response of the participant.

An MI expert collaborator developed the conversation script to move parents from the preparation stage to the later action or maintenance stages outlined in the TTM [57-59]. The MI technique provides the conversational framework to achieve the goals of the TTM and drive conversations with MICA according to the needs and goals of the parents for eating healthy, thus giving parents autonomy over their decisions [13,17]. MICA uses the script to help parents serving as proxies for their children to become motivated to adopt healthier eating habits themselves. Parents can then inculcate the same habits in their children. More specifically, the script included six commonly identified eating problems (overeating, sweetened beverages, dining out, eating fast food, stressful eating, and out-of-control eating) for parent participants to identify and self-reflect based on the existing literature [60-65]. After problem identification (agenda setting), MICA guides participants to self-appraise their perceived importance and confidence in behavior change. MICA helps participants explore previous experiences in eating behavioral changes (both strengths and barriers) by asking various questions to provide opportunities to develop strategies to overcome the identified problem. Finally, MICA prompts participants to set achievable short-term goals for the upcoming weeks. MICA also discusses long-term goals to keep participants on track toward behavior change. The same conversation script, with minor revisions, was used by MICA in both sessions to see if the experience of using MICA changes between sessions 1 and 2. For CAs to be effective for health behavior changes, continuous interaction of parents with CAs is required over time [66]. Studies investigating the role of CAs in health behavior changes often look at the single-time use of CAs. To address this gap, we conducted 2 sessions with an interval of 2 weeks to examine the participants' retention and possible changing perspectives about the system. The MI script was tweaked in the second session to include questions that follow-up on the progress with goals that the participants had set for themselves in the previous session. The conversation script also consisted of many open-ended questions beginning with *why* and *how* to help participants reflect on their habits. A part of the MI script built

by our MI expert to show how MICA interacts with participants is as follows (from a list of challenges MICA presented, a participant identifies that eating out is the biggest challenge for eating healthy):

- MICA: “Thank you for sharing. It can definitely be hard to overcome a concern like this. On a scale of 0 to 10, how confident are you in making the change to eating out too much?”
- Participant: “5.”
- MICA: “It sounds like you are uncertain of your ability to change. Why not a higher number?”
- Participant: “I think time is always a concern. We don’t have time to go grocery shopping and cook every day.”
- MICA: “Thank you for sharing. What would need to happen for you to move to a higher number?”
- Participant: “I think I can at least try [cooking] once a week.”
- MICA: “That sounds like a plan. Any other plans you can think of?”
- Participant: “I think I can start planning the whole week and find time to go weekly grocery shopping.”
- MICA: “I would love to talk about how we can address this further.”

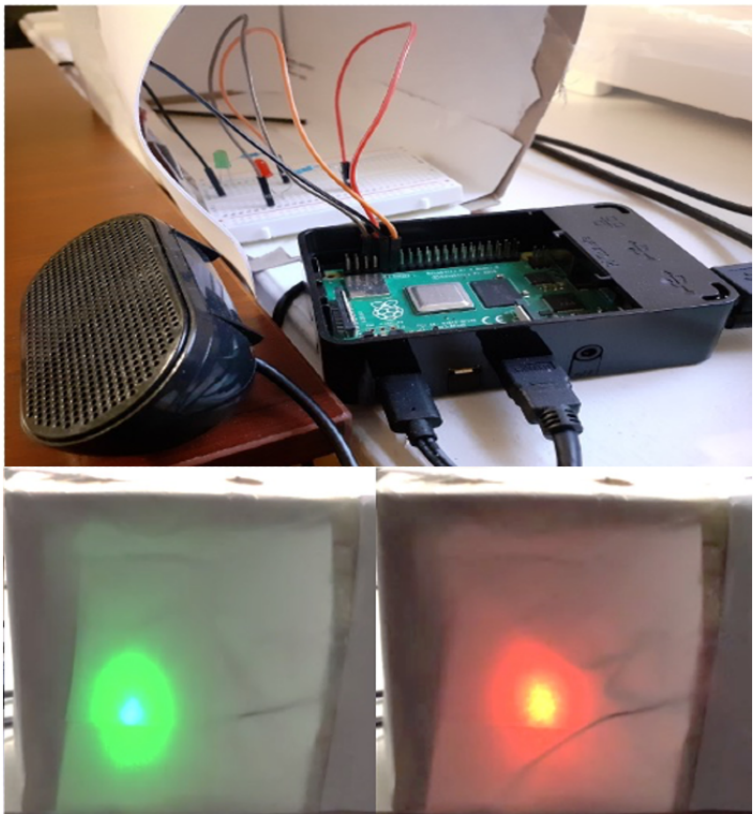
Figure 1. An excerpt from the Motivational Interviewing Conversational Agent’s (MICA) Motivational Interviewing (MI)–based conversation script. The logic tree structure of the script shows MICA’s responses to participants’ inputs of yes, no, and open-ended responses.



The MICA Prototype

The current MICA prototype runs on a Raspberry Pi platform, a computing kit manufactured by the Raspberry Pi Foundation to build computing technologies [67]. We used Python [68] to program MICA's conversation script in a JSON [69] format. Instead of relying on a third-party company's application programming interface, for example, Google Dialogflow [70] and Amazon Alexa [71], we built an in-house prototype to handle all data locally protected within the Drexel University network. The MICA prototype runs without any connection to the internet or engagement of third-party software that collects participants' input. We decided to test the limits of how far the agent can perform without cloud-based high-performance computing to accommodate the privacy concerns of potential future parent participants and those in underserved settings with limited internet access. Figure 2 shows the MICA prototype's cross-section and setup as shown to the participants.

Figure 2. The top image shows the cross-section of the Motivational Interviewing Conversational Agent (MICA) prototype containing the Raspberry Pi, the breadboard with the LED lights and wires, and the speaker. The images on the bottom show the setup of MICA working prototype from the participant's point of view used for the videoconferencing-based user study. The LED lights generate green and red lights. The green light indicates that MICA is ready to hear the participant speaking, and the red light is for when MICA is speaking to the participant.



Data Collection

As MICA is designed to be used repeatedly over time, rather than only once, we designed the study to test potential changes in participants' experiences in using the MICA for the second time. Owing to the COVID-19 pandemic [72], we conducted the study on the web via the Zoom videoconferencing service [73]. Each participant was asked to participate in 2 sessions via Zoom software [73], with a 2-week gap between the sessions (standard for MI sessions); however, the second session was optional. Participants were recruited through a screening survey posted on Philadelphia Craigslist [74]. In each session, beginning with an informed consent process, the participants filled out a pretest questionnaire containing questions on demographic information. The participants then started their interaction with MICA prototype, a 20-minute MI session starting with barrier identification to healthy eating and ending with goal-setting. The participants interacted directly with the MICA prototype through Zoom. After participants' interaction with MICA, the researcher interviewed them about the points in their interaction and communication with MICA they would like to see changed and how they envision their children using MICA. The interview also consisted of questions on how often they would be willing to use MICA and 5-point Likert scales for the perceived usefulness of MICA in reflecting on previous eating habits and motivating them to change unhealthy eating habits. This Likert scale was adapted from the System Usability Scale [75], which uses a 5-point scale for capturing sentiments about a product and is a standard benchmarking scale used in user experience studies. This process was repeated in the second session to analyze whether the experiences and perceptions changed when they interacted with MICA for the second time. This way, we would test user experiences of the system beyond participants' first impressions of the system, which is more common in user tests of CA in their early phases. The MICA testing session and the interview on Zoom were screen recorded and transcribed. A US $25 Amazon gift card was given to each participant as compensation for participating in each study session. Figure 3 shows a participant interacting with the researcher and the MICA prototype during the study.

Figure 3. The study setup for the videoconferencing-based user study. In each video conferencing session, the participant and the researcher logged in with their cameras turned on. The researcher started the videoconferencing session by showing themselves in the video (top image) and then showed the Motivational Interviewing Conversational Agent (MICA) prototype (bottom image) on the screen only when the user study began. The researcher then turned the camera back to them (top image) for the end of the study interview.

Participants

Our study was motivated by understanding the perceived usefulness and acceptability of MICA by parents serving as proxies for their children in healthy eating. Accordingly, we only conducted a study with the parents for this phase of the project. Our selection criteria included parents aged >18 years living in the Philadelphia area who were overweight or obese with a BMI [76] ≥25 kg/m^2, had at least one child aged <18 years, were not cognitively impaired, could read and write English, and were interested in eating healthier together with the child (answered "yes" for the question "Are you interested in participating in a program that addresses healthy eating behavior with your child?"). We chose parents' overweight or obese status as an inclusion criterion because it is a contributing factor to children's overweight or obesity [77]. By targeting overweight or obese parents as proxies, MICA can further impact childhood obesity.

Of the 24 parents who signed up for the first session, 20 (83%) participated in the second session. The average age of the 24 parents was 41.08 (SD 10.53) years, of which 15 (63%) were female and 9 (37%) were male. The average BMI of the participants was 31.13 (SD 5.76) kg/m^2. Each participant also provided information about their child aged <18 years and with whom the parent participant wanted to eat healthier together as a family. Table 1 shows the demographics of the parent participants.

Table 1. Participants' demographics.

ID	Age (years)	Sex	Race	BMI (kg/m^2)	Marital status	Child's sex	Child's age (years)	Present in session 2
P1	34	Female	Black or African American	25.1	Never married	Female	4	No
P2	37	Male	Asian	25.7	Divorced	Female	6	Yes
P3	35	Female	Other	42.1	Engaged	Female	6	Yes
P4	33	Male	White	27	Married	Female	5 months	Yes
P5	33	Female	White	31.9	Married	Female	7	No
P6	46	Female	White	28.3	Engaged	Male	8	Yes
P7	56	Female	White	45.2	Widowed	Male	11	Yes
P8	52	Female	White	35.9	Married	Male	17	Yes
P9	50	Female	Black or African American	27.8	Married	Female	16	Yes
P10	51	Female	White	27.2	Married	Female	8	Yes
P11	34	Female	White	26.5	Legally separated	Female	3	Yes
P12	47	Male	White	34.5	Married	Male	11	Yes
P13	45	Female	White	26.6	Married	Female	14	Yes
P14	48	Female	White	25	Married	Female	11	Yes
P15	27	Male	White	29	Never married	Male	5	No
P16	34	Male	White	34.6	Married	Female	9	Yes
P17	37	Female	Black or African American	26.6	Married	Male	5	Yes
P18	27	Female	White	36.3	Divorced	Male	2	Yes
P19	70	Male	White	30.1	Married	Male	12	Yes
P20	37	Male	White	29	Married	Male	5	Yes
P21	50	Female	White	36.7	Divorced	Male	10	Yes
P22	29	Male	White	25.8	Never married	Female	6	Yes
P23	32	Male	Asian	30.1	Married	Female	7	Yes
P24	42	Female	Black or African American	40.2	Never married	Male	17	No

Data Analysis

We transcribed the interviews and conducted a thematic analysis based on the grounded theory by Strauss and Corbin [78]. We used a constructivist framework to identify the participants' perceptions of MICA and their suggestions for improving it to help them change their own and their children's eating behaviors. The first author conducted an initial open coding on randomly selected 4 interview transcripts (approximately 10% of the total transcripts), generated codes, and presented them to the project team. On agreement and discussion from the team, the first author then continued to analyze the 40 transcripts from the 2 sessions based on the agreed codes. The analysis was done using NVivo 12 [79]. The research team met weekly to discuss newly emerged codes and themes to maintain agreement as a team. The coding structure was refined throughout the iterative cycles of discussions. After completion of the coding process, the team conducted affinity mapping [80] to identify broader themes. The team also analyzed the results of the questions on 5-point Likert scales and frequencies using descriptive statistics. As discussed in the Findings section, these measures provide a quantitative assessment of MICA's perceived usefulness between sessions 1 and 2.

Ethics Approval

This study was approved by Drexel University's Institutional Review Board (1906007221-A004).

Results

Overview

From the interviews gathered in both sessions, the participants shared how a MICA could help them to self-reflect and gain motivation for eating healthy and, as a result, be a positive influence on their children. Participants shared ways in which MICA can support them to help their children eat healthy while ensuring safety as a priority and being mindful of the complex social dynamics around children's healthy eating behaviors. The participants also shared how MICA could improve interactions to be more effective and usable.

MICA Supporting the Proxy Role of Parents

Overview

The participants felt MICA could help them self-reflect on their previous habits to make changes to their eating habits. This perception remained consistent between the 2 sessions. The participants also perceived MICA as motivational, as it helped them feel supported and motivated their eating goals amid their parenting burden. MICA could then help parents positively influence their children's eating behaviors.

MICA as an External Support Could Help Parents Self-reflect and Gain Motivation in Playing a Proxy Role

All participants reported being either familiar with or having experience in interacting with a CA, such as Google Home, Amazon Alexa, Samsung Bixby, or Apple Siri. For the participants, the difference between MICA and other CAs was that MICA asked healthy eating–related questions that made them reflect on their previous eating habits (P4, P14, P16, and P21). This self-reflection could help them influence their children's eating behaviors as a proxy. The participants also saw MICA as a "close friend that motivated and nudged [them] to eat healthier" (P8, P11, P16, P19, and P23). When asked to choose 1 eating problem they wanted to solve as a parent who may be role modeling the eating habits of their children, participants chose eating junk food (5/24, 21%), overeating (4/24, 17%), emotional overeating (3/24, 13%), nighttime snacking (4/24, 17%), eating too much sugar (3/24, 13%), stress eating (2/24, 8%), eating fast food a lot (2/24, 8%), and the ability to self-control their eating (1/24, 4%).

As the participants would often get busy taking care of their children and helping them eat healthy, having external support in MICA was perceived as helpful. MICA could help participants understand where they were going wrong with their eating habits and what needed to be changed. By using MI in its script, MICA was able to gauge deeper into how the participants' eating habits related to the family's eating habits. The MI script allowed the participants to understand the current state of the family's eating habits before thinking about how to change their own eating habits. Phrases such as "I see, tell me more" helped participants to reflect deep into the response they gave to MICA:

> *I like that if she [MICA] asked me a question and I did answer–It was as if it [MICA] understood my response and it [MICA] went back and it [MICA] would ask me another follow-up question based off of that answer.* [P15]

> *But I remember when I was talking about the energy–I kept my answer very short, and then the follow-up question was, "Can you please tell me a bit more about that." So, it shows that she [MICA] will go the extra mile.* [P11]

Following up on a question made the participants (7/24, 29%) feel heard and cared about, a need that often gets ignored for parents serving as a proxy for their children. Overall, 42% (10/24) of participants especially liked the question, "If you were to stick with your current plan to change, where do you see yourself in 5 years?" as it made them think about the long-term effects of changing their family's eating habits. This question also helped the participants prioritize activities for themselves and their children that could lead to a long-term behavior change. In all, 38% (9/24) of participants also expressed that the process of conversing with MICA made them verbalize their actions and kept them "accountable" for their eating goals. Being accountable to achieve their goals becomes essential for parents serving as a proxy for their children who want to bring about the same change in their children.

On the basis of participants' perception of MICA as a tool for self-reflection, we followed up with a Likert-scale question to gauge the effectiveness of MICA. For the question of *whether MICA helps to reflect on previous eating habits* on a 5-point Likert scale of 0 being "not helpful" and 5 being "very helpful," the median score was 4 (IQR 3-5) in session 1, leaning toward helpful or very helpful. For session 2, the median score was 4 (IQR 4-4). The participants explained this perception by sharing that verbally conversing with MICA pushed them to think more about their previous eating decisions:

> *[MICA] just kind of helps bring to light what's been happening forever, and just kind of realizing what I was doing was a very helpful thing, you know. If I am not made aware of things, then I probably won't stick to my goals. You know, I just kind of fell into a slump, and so reacting to MICA kind of helps realize what I was doing. So, it was very helpful.* [P4]

Furthermore, 50% (12/24) of participants perceived MICA as "motivational" because of the positive reinforcement (9/24, 38%) and affirmations (6/24, 25%). The MICA being "motivational" could help participants adhere to their healthy eating goals, especially when their self-motivation was low. This problem frequently happened during their caregiving process as a proxy, where parents often felt that they had no other support:

> *But having somebody speak to, motivational, you know, I guess it motivated me to want to do this. I can do this.* [P8]

> *It would leave me feeling like, "Okay, I think I got this. MICA thinks I can do this. I think I can do this, even if nobody else in my life thinks I can," you know, there is something backing me up.* [P16]

On the basis of the participants' perception of MICA as motivational, we followed up with a Likert-scale question. For the scale measuring *whether MICA motivated participants to change eating habits*, the median scores were 4 (IQR 3-5) and 4 (IQR 4-5) for sessions 1 and 2, respectively. MICA's MI script used empathetic phrases such as "I can understand how difficult it can be to bring about a change such as this" to empathize with participants' problems as proxies and motivate them to overcome those problems:

> *But it doesn't sound as though [MICA] is just reading a script. It's like there was a lot more thought put into the process, and I will definitely say a lot more empathy also. It's very much appreciated because it goes to show that even though [MICA] is, you know,*

> *not a human being that they [MICA] can still feel for you and your struggle.* [P3]

The results indicated the acceptability of MICA as a tool for helping the participants reflect on their eating habits that affect their children, despite their busy lives as parents serving as proxies. MICA was also perceived to provide mental support to the participants and motivate them to eat healthy. As MICA helps participants to eat healthy, they will be better motivated and supported to bring about the same behavior change in their children.

MICA Could Help Parents to Positively Influence Their Children as a Proxy

Currently, MICA focuses on helping parents change their own eating habits. As a proxy for behavior change, parents are then motivated to bring about the same behavior change in their children. Participants shared that MICA could help them replicate their healthy eating habits in their children. To better achieve this goal, the participants wanted MICA to provide them with informational support so that they could independently form strategies with their children for eating healthy.

A challenge that 29% (7/24) of participants faced in their everyday lives was that although they encouraged their children to eat healthy, the participants themselves did not eat healthy. Being their children's primary caregivers and having trouble adopting healthy eating habits of their own, the participants found it challenging to incorporate healthy food items in their children's diets. This problem inhibited the parents from becoming good role models or proxies for their children to help them adopt healthy eating habits:

> *I [parent] am suffering to get her [child] to eat the greens and veggies or whatever. But if I don't eat it, then who am I to force something on her if I don't make it for myself or like it. So, it has been a challenge.* [P11]

To address this problem, the participants perceived that incorporating healthy eating habits themselves with the help of MICA would motivate their children to eat healthy with them:

> *Him watching me interact with [MICA] and eating better, he [child] may, you know, want to do the same thing.* [P20]

> *Well, I feel like if I'm eating healthier, and I am pretty much the main cook in the house. Then I'm going to incorporate healthier items into his [child's] plate.* [P15]

MICA's current design following the MI principles probes participants to develop healthy eating strategies of their own. For young children to best respond to these MI strategies through their parents, the participants wanted to be involved in their children's food decisions, with the aim of their children gaining greater control over their food decisions gradually over time. They wanted MICA to offer suggestions on healthy eating, such as the food items they could incorporate as part of healthy eating (P7, P15, P21, and P23). Nutritional information about the different food items was another suggestion from MICA that 33% (8/24) of participants wanted to make their children aware of the food items they consumed. Apart from the motivation that MICA provides, the participants believed that these suggestions and information would help them prepare their children to make informed food choices over time and gain more autonomy over their eating habits:

> *And, you know, [parent] can help them obtain really good habits when they're young. I [parent] used to have really good habits because my mom would teach me about the kind of the food groups and how those things react. It's those changes in my adult life that have gotten me to where I am. So, I would just think would be a great way for kids to get more information and make better choices for themselves.* [P7]

Parenting, however, is often not an easy task to be undertaken by 1 guardian. Multiple family members and households are often involved in influencing children's behavior.

MICA Safely Working With Complex Family Dynamics and Trust

Overview

One of the challenges participants wanted to resolve about their children's eating habits was managing conflicting expectations between parents, guardians, and family members. They felt MICA could work as third-party support to help them navigate this complex dynamic. Furthermore, participants wanted MICA's conversational approach to address their concerns around existing health behavior change tools by putting safety before behavior change goals.

MICA Should Understand and Work With Complex Family Dynamics as Third-Party Support

Participants felt MICA could help them keep track of their children's eating habits when dealing with complex family dynamics, such as managing children between multiple households and differing parenting styles among guardians. Participants felt it would be useful if MICA plays a third-party role, such as providing them suggestions and reminders to help their children understand the importance of eating healthy.

P16, P21, and P22 were concerned that their children were developing unhealthy eating habits or gaining a negative influence when visiting the other parent from whom they were separated. This finding also portrayed the complex social dynamics involved in families' healthy eating:

> *Our nine-year-old talked to my wife that she was feeling fat. And we share custody with my ex, and my ex had an eating disorder. My ex's mother actually tells my daughter that she's too fat.* [P16]

Similarly, P3 felt that she did not have the proper authority to teach her partner's child to eat healthier. Such social dynamics involved in families' diverse social relationship settings make it difficult for participants to take complete control of healthy eating practices. Some participants (9/24, 38%) were also single parents or separated from their partners. In such cases, MICA could become a "support system" to help the participants keep track of the child's eating habits and provide a consistent eating regime for the child even while visiting the other parent:

> *[MICA] provid[ing] affirmations or reminders to the other parent is super helpful because when she's [child] away from us, she [child] doesn't have a huge social support system to combat those negative influences. So, for her, even if MICA just says to the other parent, "Hey, remember to eat veggies and make sure to get enough protein and starches," then, in that case, MICA is providing a great influence that she [child] otherwise wouldn't get, um, other than that.* [P16]

Some participants (7/24, 29%) also expressed difficulty conveying the importance of eating healthy to their children:

> *I still find it difficult for me to explain to my child about the benefits of healthy eating.* [P23]

The participants (7/24, 29%) had a tough time making their children "listen" to them, as their children ignored their suggestions for eating healthy food that they did not like to eat:

> *99% of the time they [children] don't listen. They're [children] not open to trying new things.* [P20]

To help with this challenge, P13, P14, P16, P14, and P22 envisioned using MICA to remind their children of the benefits and ways of eating healthy. P22 suggested having MICA help them to "remind [child] to eat vegetables and giving [child] information and why it's important to eat vegetables and healthy food."

MICA Should Foster Trust and Safety by "Figuring Things Out Together" and Involving Experts

The participants were concerned that pushing children to eat healthier can potentially generate side effects such as undereating. P4, P16, and P19 saw undereating because of such peer pressure as a challenge that MICA could potentially address; they thought MICA could approach healthy eating differently than other interventions in ways parents and children can talk about it together and "figure out things":

> *I would hope that she [child] could ask questions, like, "How much should I've eaten" or "What did I eat." Or, not make [eating healthy] turn into a thing to where she's so focused that she developed an eating disorder. But, something that's more like a tool [MICA] where, you know, we can talk about eating together and figure out different things.* [P4]

The participants felt that MICA could support them to help their children in the longer run by preparing them to make informed decisions about their eating habits. However, the participants did not completely trust MICA's suggestions for helping them with their children. Especially when the children were younger, they were concerned about the misuse of CAs to aid in undereating or overeating (P11, P12, P18, and P19). P16 and P19 also desired the involvement of a health care provider or at least some parental supervision over how MICA incorporates suggestions for their children because they did not trust that MICA could handle sensitive complexities around pressure to eat healthier, potentially leading to undereating:

> *I don't know that without like a physician involved, I wouldn't really want to use it for anything more than just generic nutrition advice. Like, what's healthy about an apple– I would completely trust MICA to tell me that. But, if my daughter decided to come up with a weight loss plan and talked to me about it, I would be really, really concerned about the ability of a voice assistant to help me handle the complexities of that.* [P16]

Here, we discovered that the participants wanted MICA to help them guide their children on healthy eating. However, enabling CAs as a guide for helping parents with their children can be harmful to children, as CAs possess the risk of being misused by external parties without parents' knowledge. CAs can then generate inaccurate advice and suggestions, resulting in more harm than good. Parents may not be informed of such harmful interactions, making it difficult to keep track of the harm children may be going through.

Enhancing the Interaction Experience With MICA

Overview

The participants perceived that MICA's motivational and empathetic responses could make them feel supported in their parenting process. MICA was then compared with a therapist or a nutritionist because of the accountability and empathy it provided. Because of this perception, participants shared how MICA's interaction can be improved for a better therapeutic experience. One of the major themes that emerged was whether to accommodate human-like interactions over non–human-like interactions. Participants also wanted MICA to enhance goal-relevant interactions in its MI scripts. Finally, how often, where, and how long they wanted to interact with MICA also mattered to the participants.

MICA Should Customize Human-Like Versus Non–Human-Like Interactions to Parents' Preferences

A recurring theme in improving MICA was to make it more or less human-like. Being more human would involve making MICA's conversations to be more "natural" (7/24, 29%) and the voice to be less "robotic" (15/24, 63%). MICA's human-like nature would help participants feel less lonely, serving as a proxy for behavior change in their children. On the other hand, some participants (8/24, 33%) preferred that MICA was "artificial" so that they could talk about their and their children's eating problems in a "safe manner to an AI without being judged."

The participants often used human pronouns to refer to MICA, such as she or he. They felt that MICA ingrained components of a human-like conversation, as it acknowledged their previous responses, indicating that MICA was following up on their responses. This human-like nature of MICA could help the participants feel supported and less lonely in their behavior change journey along with their day-to-day caregiving tasks. MICA also referred to previous answers given by the participants both during a session and between the first and second sessions. The participants liked this call back to their previous conversations, as it meant to them that there was somebody (MICA) to check up on them as well, a task which they did daily for their children:

> *I like that it [MICA] was able to recognize what you're saying and then be able to provide thoughtful follow-up questions for it.* [P2]
>
> *She [MICA] remembered what I did when I said my problem was last week, and now she's checking on me to see how I'm doing with that. How nice obtaining that information!* [P14]

For improving the human-like interaction further, the participants wanted longer, more drawn-out conversations that look into "the reasons behind an answer" (P12, P14, P15, and P22) for behavior change, as people often try to get to the bottom of an answer in human-human conversations:

> *To give it more of a feel like I'm actually talking to a person– have conversations where I can expand on a thought and get a little bit more detailed– where I feel like she's [MICA] going to understand me.* [P14]
>
> *I would say probably going more in-depth to each aspect. So, it [MICA] kind of digs into where you're at, what your plans are, and things like that.* [P15]

On achieving a goal the participants had set for themselves, MICA responded with positive phrases such as "That's awesome!" or "I am glad to hear that." The participants perceived such interactions could motivate them to continue with their goals and offer mental support that is hard to receive as parents serving as a proxy. However, P9 found it "awkward" to hear words of encouragement from an artificial device. She preferred more direct responses from MICA, wherein no affirmations or positive words such as "That is great to hear" or "Awesome!" were used. P21 also found positive encouragement from MICA "not as sincere as a human" as she perceived it not to be grounded in "emotional reality." In contrast to making MICA human-like, P7, P12, P18, P20, and P21 preferred MICA not to be human-like, as it meant it would not "judge" them or "get mad":

> *I mean, honestly, it's probably easier for me to answer a question from an AI about my eating habits or what I think I'm doing wrong or something like that than it would be to answer questions like that from a real person. Because I don't feel like the AI is going to judge.* [P21]

MICA not being human-like made it "trustworthy" (P11 and P23) and "comfortable for sharing things like weight" (P18) for some participants in the context of accomplishing their eating goals, especially information related to their children.

> *It lowers that apprehension that you have with things or people or whatever and immediately ups the trust that you have with them.* [P11]

Participants especially preferred MICA being non–human-like in the context of adopting healthier eating habits, which can become a sensitive or personal topic to share with a human-like partner.

Participants liked that MICA followed up on their responses and suggested ways to make it more human-like. On the other hand, 29% (7/24) of participants preferred MICA to be non–human-like. This finding calls for giving participants the option to change MICA according to their preferences, making them feel supported in behavior change while providing sufficient help as a proxy for their children to eat healthier.

MICA's MI Script Should Reflect Customized Goal-Relevant Questions and Suggestions for Parents

MICA helping participants achieve their own goals for eating healthy would help them replicate the same behavior change in their children. In this regard, 38% (9/24) of participants shared the kinds of questions they prefer MICA ask to help them make better decisions for themselves and their children. The questions asked by MICA as part of its MI script were thought to be relevant and "thought provoking."

Some participants (10/24, 42%) enjoyed the time-based question as part of the goal-setting agenda, wherein MICA asked them if they stuck to their goals, where they would see themselves in 5 years. This form of time-based goal setting could help put things into perspective and provide the necessary motivation for participants to change their and their children's eating habits. However, P5, P14, P15, and P17 also wanted MICA to change the period of these questions to shorter durations of a week, a month, or a year. They perceived that being able to change this period could provide them the necessary motivation to achieve short-term goals:

> *And, maybe, set a weekly goal amount. As the weeks go by, you know, with weight loss, in the beginning, you lose a more significant amount, and then after that it starts to go a pound here and a pound there, you know. So, maybe that could also be taken into consideration as time goes on. Because, the way she [MICA] asked the question, I kind of felt it's long-term because she said that about five years from now.* [P14]

As parents shape their children's eating behaviors from an early age, having questions that reflect the problems and needs of parents would be essential for supporting the participants as proxies. For improving MICA to be reflective of their goals and needs, 38% (9/24) of participants wanted MICA to ask relevant questions first to gauge their daily dietary behaviors:

> *It didn't get specific as to the foods I'm eating. Do I eat a lot of salty snacks? Am I more loading up on carbs? It [MICA] really didn't ask me all those kinds of questions that much.* [P13]

Similarly, P4 desired more questions on his problem of nighttime snacking, asking about what food items he consumed late at night and what he would do to overcome this problem. P15 and P22 wanted MICA to ask about the food items they liked to eat and "couldn't resist." P9, P10, P11, P20, and P22 wanted more questions on what food items they were consuming every day and what they would and would not like to eat as a part of healthy eating. The participants then wanted MICA to adapt to their individual goals and needs by asking them questions about their plans for accomplishing their goals:

> *[MICA] can ask questions such as "What are you planning on having today for breakfast?" or just maybe like mix it up every once in a while, and say, "Hey, what's your plan for today with your eating goals?" Or, maybe, suggest to me, "Hey, have you ridden your bike today?" or, "Have you gone for a*

> *walk today?" [MICA] gives helpful little tips about keeping healthy and staying healthy.* [P10]
>
> *Questions will be centered around like "How are you doing with this?" or "What have you eaten today?" and those types of questions, I think, could enhance the personal aspect of the interaction with [MICA].* [P20]

This interaction based on the participants' dietary choices and problems could help them make informed decisions about which food items to eat and which to avoid, helping to replicate the same for their children. Another way some participants (7/24, 29%) wanted to improve their interaction with MICA was by MICA making its suggestions specific to their goals. P8 wanted to be able to decide what to eat for a meal based on what MICA marked as healthy, for instance. P23 suggested MICA inquire about the "types of food consumed" during a day to recommend coping strategies; P15 asked for "sharing information about like one soda has this many calories"; and P10 asked for MICA to keep track of the sugar-based foods consumed during the day given their prediabetic condition. Such information provided by MICA would help participants take better care of themselves and, in turn, their children.

Location, Frequency, and Duration of Interacting With MICA Matters

When we probed the participants on how they might envision using MICA in their home contexts, the initial reaction was where they would place it. The meanings that the participants associated with the physical space about healthy eating and how MICA could best support them in their daily caregiving tasks mattered in their perceived choice of the physical placement of MICA. The participants wanted to place MICA in the kitchen (16/24, 67%), bedroom (6/24, 25%), living room (1/24, 4%), or working desk (1/24, 4%). This perception remained consistent between the 2 sessions. The kitchen was the place in the house where the participants made most of their healthy eating decisions together with their children. Participants shared that having MICA there as a reminder for healthy eating could help the participants not make "unhealthy" decisions for themselves as well as their children and could keep them accountable:

> *[Kitchen] is where the offending actions have been. So, if she [MICA] was there [in the kitchen] as a friendly reminder, you know, you've got goals so be accountable to them.* [P7]
>
> *So, you know, if it's in another room versus the culprit room we'll call it where most of the unhealthy eating happens, which is like the kitchen area. If MICA is there [in the kitchen], I'll be more inclined to have her perked up by, you know, addressing her and the conversation will start.* [P3]

The perceived frequency with which the participants saw themselves conversing with MICA decreased from sessions 1 to 2, as they realized the feasibility of taking time out from their busy caregiving schedules to converse with MICA. For session 1, of the 24 participants, 18 (75%) participants reported wanting to interact with MICA daily, 5 (21%) participants every week, and 1 (4%) participant did not report. For session 2, of the 20 participants who followed up, 9 (38%) participants wanted daily interaction, 1 (4%) wanted interaction with the MICA every 3 to 4 days, 6 (25%) wanted it to be weekly, 2 (8%) preferred biweekly, 1 (4%) participant reported monthly interaction, and 1 (4%) occasionally when desired. P16, P19, P20, P23, P4, and P8 wanted daily interactions with MICA, as it meant receiving motivation to adhere to their goals regularly. After the second session, P2, P3, P18, and P21 felt that less frequent interactions would give them more time to help their children while making sufficient progress toward their goals, which they could then report back to MICA. P21 elaborated on her preferred frequency of a week as follows:

> *Every day would be too much. But over the course of a week, you've had highs and you've had lows. And so, I think that would be a better self-reflective period.* [P21]

P4, P13, P21, P23, and P24 also said they would prefer less frequent interactions with MICA as they often got busy providing care and would find it challenging to take time out to converse with MICA. Participants (9/24, 38%) with similar opinions on frequency still wanted a "daily reminder system," which would softly nudge them toward their goal:

> *So, I would want [MICA] to do the reminders every day. Not in an accusatory way, but like, "Hey, we're able to go take your walk today." And that really might make me be like, yeah, I need to go do that.* [P8]

The participants (9/24, 38%) also felt they often did not have enough time to commit to a uniform session length with MICA. As the participants would get busy taking care of their children while trying to change their own and their children's behavior, the conversation length could vary depending on the time the participants had for their interactive sessions with MICA. The participants perceived that having the current MICA session followed by smaller sessions based on the participants' availability would be feasible for the participants:

> *I'm wondering if you had many sessions with MICA–this session is an introductory session and then there are some follow-ups, you know, just checking in or that kind of thing.* [P7]

This change in perception for how often and how long they want to interact with MICA after the second session highlights the need for MICA to cater to the changing needs and preferences of parents serving as a proxy for their children. Participants only knew what felt right after having experienced it. Using MICA for the second time in 2 weeks made them realize they needed more time to achieve the goal they had stated they wanted to achieve during the first session. For parents serving as a proxy, whose needs and goals mutually affect the needs and goals of their children, MICA should give parents the option to change the frequency and length of the sessions as desired by the parent.

The results report on MICA's perceived usefulness in helping participants reflect on their eating habits and stay motivated toward their healthy eating goals, which indirectly helps their children to adopt healthy eating habits. Participants perceived that it mattered where MICA is placed or how often they interact

with MICA and for how long. These configurations would then be adjusted based on each participant's family's unique goals and needs around accountability and motivation.

The results show the different roles and functions of an MI-based CA that can help parents eat healthy with their children by serving as proxies. The findings point toward the design of CAs that go beyond supporting a target individual of behavior change and supporting behaviors that require coordination between multiple persons, such as between parents and their children.

Discussion

Overview

Our research was primarily motivated by the lack of accessible technological support for parents, who often act as proxies for their children for behavior change. We conducted a study to test user experience around the concept of an MI-based CA, MICA, that supports parents to help them eat healthier together with their children. We learned several lessons, starting with requirements for supporting behavior change that requires coordination between parents and children in various social complexities. We learned how MICA's interactions could be customized to cater to the goals and needs of parents because they serve as proxies for their children's healthy eating goals. We also learned how MI-based conversations can be incorporated in CAs to cater to the needs and goals of parents serving as proxies. We end with discussing broader implications for designing MI-based CAs for those serving as proxies for health behavior change.

Principal Findings

On MI-Based CAs Supporting Parents to Safely Manage Children's Behaviors in the Context of Healthy Eating

Existing research suggests using CAs to track children's food intake and then nudging the children to eat healthier under parents' rules [2]. The literature also indicates parents' apprehension about their children interacting directly with a CA [2,41]. Currently, MICA implicitly helps parents serve as proxies for their children in eating healthy. MICA provides support to parents to change their eating behaviors, who can then motivate their children to eat healthy. Parent participants in our study suggested acting as role models for their children, wherein their children could replicate their eating habits or learn from them as they made progress with CAs. This proxy relationship can be beneficial for both parents and children, as they motivate each other to eat better [81]. In addition to implicit support, our findings show that parents wanted MICA to provide explicit support and information for help with their children. For example, MICA can suggest strategies to parents around conversations with their children about their eating habits and come up with goals together as a unit.

Diverse family structures (eg, parenting partner's child, one's child spending time at another parent's household) in our participant pool generated complex social dynamics that the participants thought MICA should address. Studies have discussed CAs serving as social actors to mediate communication between family members [31,82,83]. From our findings, MICA could help parents provide consistency in the changing lives of their children for maintaining a behavior change. This consistency could be brought about by MICA serving as a uniform guide for the child toward making healthy food decisions, even when they were with the other parent who might have different approaches to healthy eating. Thus, CAs can help parents change their children's behavior even in their temporary absence.

The participants expressed that MICA's MI-based approach can contribute to strategies around helping parents support children to gain autonomy over their eating decisions from an early age. Although a parent may be comfortable involving a CA for guidance, CAs always have the potential of being misused. As parents develop trust in the CA through continued use, external parties can use the CA to manipulate the information given to the parents and therefore to the children. Moreover, language models are often not competent enough to interact in a safe and factually correct language consistently [84]. To find a balance between child safety and CAs supporting parenting, parents should be given appropriate background knowledge about CAs' suggestions and recommendations for healthy eating strategies. Such transparency from CA designers about the trustworthiness of such technologies is a good first step toward developing CAs for parenting [85]. In this regard, regulating the suggestions given by MICA to parents acting as proxies for their children becomes an essential feature. Involving a health care provider or an MI expert to regulate MICA's interaction would be helpful, similar to how the current MICA is designed. Every response of MICA is strictly designed by an MI expert. This finding highlights the need to make it explicit to parents that MICA is strictly guided by an MI or health expert; explicitly stating this information to parents can increase trust among them for similar CAs.

Undereating or overeating due to peer pressure was a prevalent problem among the children of the participants in our study. To overcome this problem in their children, participants suggested MICA help them in monitoring inaccurate advice that their children may receive through other external channels, such as peers or social media [86]. Therefore, in addition to ensuring that the content's credibility is high, finding ways to support parents to moderate CAs' content will be helpful. However, given the largely black-boxed algorithms of artificial intelligence–based CAs, moderating CA's responses from parents' perspectives will be a critical design challenge.

On Customizing Interactions of MI-Based CAs for Parents as Proxies

Customized conversational strategies in CAs improve system performance, usability, and efficiency [87-89]. Parents, who are the informal caregivers of their children, often have the responsibility of making informed decisions on behalf of the children. Findings from the participants supported the need to include customizable features to meet the varying needs of parents serving as proxies. For example, the parent participants wanted to change their conversation length with MICA depending on how much time they had and their personal preferences. Participants thought that the current MICA session

length followed by smaller sessions based on their availability would be feasible for them. Parents acting as proxies for the persons they care for are often overburdened physically and mentally [23]. Therefore, it becomes necessary for these proxies to adjust certain CA features depending on their availability and use. Studies have also shown that adapting conversation length based on user preferences may lead to fewer concerns about content credibility and may increase trust in the system [90]. Other standard features that CAs could adapt to proxies' preferences are the agent's verbiage, sex, accent, language, and volume.

The meaning that the participants associated with the physical space for healthy eating affected the physical placement of MICA. MICA's placement can be customized to the needs of the participants and how MICA can best support them in their daily caregiving tasks. A total of 67% (16/24) of participants wanted to place MICA in the kitchen, as the kitchen was the place in the house where the participants made most of their healthy eating decisions together with their children. The frequency of interaction with MICA also differed between the participants, but they wanted less frequent interactions after the second session. After the second session, the participants felt that less frequent interactions would give them more time to help their children while making sufficient progress toward their goals, which they could then report back to MICA.

Owing to its empathetical conversational style, our findings showed that MICA can be seen as a human-like agent. Aligned with previous research on people's perception of CAs [91-93], empathetic responses from a CA would make it appear more like a human, thus helping to alleviate the social isolation of a parent as a proxy to some extent. Parents may call a CA by the name or a human pronoun (eg, "her") to keep them accountable and feel empathy and support from it. However, being human-like is not always perceived as favorable by parents [94], as it often can be perceived as insincere, coming from an artificial agent [95]. Empathetic and motivational responses from an artificial device can be "awkward" and "weird," as was found by 33% (8/24) of participants in our study as well. Of those parent participants, 21% (5/24), even preferred that MICA was an artificial being—it appeared to "not judge" them on their eating behaviors and made them feel "comfortable" to share their weight. This contrasting perception of MICA in our findings on what is appropriate or effective calls for the human-like nature of CAs to be modified based on the context and parents' preferences. The preference for how human-like a CA should be may change for the same parent at different times of the day, similar to how 4% (1/24) of participants preferred to hear motivational words at the start of the day compared with the rest of the day. This customizable human-like nature of CAs would make parents serving as proxies feel supported and help them focus on their caregiving tasks at hand.

On Designing MI-Based Conversation Scripts of CAs for Parents as Proxies

Parents help shape their children's eating habits from an early age, and involving parents in interventions as core participants helps improve children's healthy eating [7,20,21]. Accordingly, helping parents eat healthier would help their children eat healthier as well. Parents often find themselves overburdened with providing care, leaving them vulnerable to mental health problems and a lack of support. Through the use of MI, MICA could help the participants feel supported and motivated in their healthy eating journey. The MI technique emphasizes a personalized approach to behavior change, offering autonomy to parents over their behavioral goals and decisions. This form of self-regulation and positive motivation may result in the long-term maintenance of behavior change, especially for parents as proxies, who may require motivational support to bring about a behavior change in themselves and their children [96,97]. However, such MI technique requires a continuous conversational exchange. Parents serving as a proxy are continually asked to answer questions that would probe about their and their children's hidden challenges, motivations, and goals.

By incorporating the MI technique, MICA allows parents as proxies to develop their own goals based on their context, strength, and needs. MICA probing participants about their barriers to eating healthy led them to self-reflect on their eating problems and solutions, which were all essentially generated by the participant, not by MICA. The participants felt that having an external member point out opportunities in their behavior change regime for self-reflection could greatly help them, as they did not have the time or the support to do so themselves. Such self-reflection could motivate the parents to bring about the same behavior change in their children for whom they are proxies. The work by Chen et al [3] on parents using voice-based technologies at mealtime found that parents did not like to receive instructions from a CA, as they feared it would make them and their families dependent on the CA. Our work further adds to this finding that even with a technology-mediated approach, using MI as a conversational tool for interacting with parents will give them autonomy and motivation toward their health decisions. This form of interaction can further help parents feel more confident in helping themselves and their children.

Asking time-based questions as part of the goal-setting phase of MI for eating healthy was also preferred by the participants to stay motivated. The time-based question "Where do you see yourself in 1 year?" could be asked for goals that they could accomplish within a short period; for example, not eating out frequently. On the other hand, "Where do you see yourself in 5 years?" could be asked for goals that required more time to accomplish according to the participants; for example, the goal of losing a specific amount of weight. Such time-based goal-setting can be a catalyst for providing the necessary motivation to bring about a behavior change in parents [98]. Parents can, in turn, help their children achieve the same behavior change. The period in which the parents see themselves accomplishing a goal is also dependent on how much time they can devote to the goal from their caregiving schedules.

MI-based conversations offer CAs the opportunity to dive deep into the parents' problems, as is possible in human-human conversations. Furthermore, as MICA's MI-based conversation with the participants was not restricted to performing a task, the participants felt the conversation to be more "natural" [99]. CAs are then compared with a "therapist" or a "guide/coach,"

although more complex language models have to be developed for CAs to become truly conversational [84]. Although MICA helped participants reflect on their responses, 17% (4/24) of participants still felt that conversations with MICA could be more drawn out. MICA has the potential of going deep into the problems faced by parents and understanding where they are with a goal. Although diving deep into the parents' problems and coming up with solutions is useful, changing the frequency of such conversations based on parents' availability would increase retention of use, especially for parents who are otherwise occupied with providing care. Rather than conducting the same session every day, CAs should give daily reminders about the goals parents have already set for themselves and their children. Therefore, although MI can help parents as proxies gain confidence in their decisions, frequent MI conversations may not be deemed feasible. CAs should incorporate a balance between MI conversations and daily reminders, which must be customized according to the needs of the parents serving as proxies [90].

Toward CAs Supporting Those Serving as Proxies for Health Behavior Change

User testing parents' experience of using MICA provided a building ground for designing CAs that cater to the needs of those serving as proxies beyond parents. Parents wanting to eat healthy together with their children is one example of a proxy relationship for health behavior change. In addition to parents supporting their children, proxy relationships can include any family member or friend supporting a close one. Such proxy relationships are common when the target individual is unable to directly interact with the source of behavior change [8,9]. Examples include adult children of older adults taking care of their parents physically and emotionally or a close friend helping their friend with multiple chronic conditions needing to exercise together for medical reasons. Another example is a spouse helping a cognitively declining spouse who needs to perform daily physical rehabilitation exercises. Current technology providing solutions for health behavior change poses a limitation of not going beyond the needs of individual users and accommodating the needs of such proxy relationships. Technological interventions may not be accessible to care receivers such as young children or individuals who lack access to technology because of socioeconomic reasons, including costs. Those who lack appropriate technical literacy, health literacy, or physical capacity to use such interventions directly are also at a disadvantage. Therefore, it is essential to design technologies that cater to such proxy relationships to address these gaps in fulfilling the unmet needs of individuals and their supporters. Little is known beyond how individuals or a group of individuals directly interact with health-focused CAs. Our study provides a starting point for the critical role CAs can play in leveraging social contexts, such as proxy, where multiple people are involved in an individual's behavior change.

Limitations and Future Work

This study was limited by the context of using MICA in a laboratory setting. Furthermore, as the study was conducted remotely due to the COVID-19 pandemic, the presence of a researcher during the interaction of the participants with MICA may have influenced their responses. Future studies with an improved MICA prototype could involve testing MICA with parents in the privacy of their households. The MICA prototype was tested in 2 sessions 2 weeks apart to understand changes in the participants' perceptions and needs. To measure the long-term behavior change effects of MICA, a longitudinal study spanning weeks can be conducted in the future.

Conclusions

As off-the-shelf CAs are increasingly adopted in people's homes, opportunities to automate and increase access to behavior change techniques that benefit from conversational exchanges, such as MI, arise as a sustainable behavior change solution. Family, parents, and friends who act as a proxy for the behavior change of a target individual are often burdened with various support tasks. We developed an MI-based CA called MICA to help parents support their children better as a proxy for health behavior change. This study aimed to examine the acceptability and user experience around using MICA by parents serving as a proxy for their children in adopting healthy eating habits. From 44 user test sessions, several insights emerged around the design requirements of MI-based CAs that support multiple individuals involved in a person's behavior change. Future work will examine how MICA can lead to effective long-term behavior change. This study is the first attempt at investigating how CAs can support behavior change influenced by multiple people. Our work contributes to the increased attention on CAs for health and how CAs can leverage such a critical social context—a proxy—in supporting the health of individuals who have limited ability to take care of themselves.

Conflicts of Interest

None declared.

References

1. Kinsella B. The information says Alexa struggles with voice commerce but has 50 million devices sold – one of those seems right. Voicebot. 2018 Aug 6. URL: https://voicebot.ai/2018/08/06/the-information-says-alexa-struggles-with-voice-commerce-but-passes-50-million-devices-sold-one-of-those-seems-right/ [accessed 2022-04-19]
2. Garg R, Sengupta S. He is just like me. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-24. [doi: 10.1145/3381002]

3. Chen Y, Li Z, Rosner D, Hiniker A. Understanding parents' perspectives on mealtime technology. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-19. [doi: 10.1145/3314392]
4. Reedy J, Krebs-Smith SM. Dietary sources of energy, solid fats, and added sugars among children and adolescents in the United States. J Am Diet Assoc 2010 Oct;110(10):1477-1484 [FREE Full text] [doi: 10.1016/j.jada.2010.07.010] [Medline: 20869486]
5. Breiner J, Beck S. Parents as change agents in the management of their developmentally delayed children's noncompliant behaviors: a critical review. Appl Res Ment Retard 1984 Jan;5(2):259-278 [FREE Full text] [doi: 10.1016/s0270-3092(84)80006-5]
6. Golan M. Parents as agents of change in childhood obesity--from research to practice. Int J Pediatr Obes 2006 Jan;1(2):66-76. [doi: 10.1080/17477160600644272] [Medline: 17907317]
7. Bourcier E, Bowen DJ, Meischke H, Moinpour C. Evaluation of strategies used by family food preparers to influence healthy eating. Appetite 2003 Dec;41(3):265-272. [doi: 10.1016/s0195-6663(03)00104-1]
8. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
9. Bandura A. Toward a psychology of human agency: pathways and reflections. Perspect Psychol Sci 2018 Mar 29;13(2):130-136. [doi: 10.1177/1745691617699280] [Medline: 29592657]
10. Beshara M, Hutchinson A, Wilson C. Preparing meals under time stress. The experience of working mothers. Appetite 2010 Dec;55(3):695-700. [doi: 10.1016/j.appet.2010.10.003] [Medline: 20937335]
11. Bernas KH, Major DA. Contributors to stress resistance: testing a model of women's work-family conflict. Psychol Women Q 2016 Jun 24;24(2):170-178. [doi: 10.1111/j.1471-6402.2000.tb00198.x]
12. Neumark-Sztainer D, MacLehose R, Loth K, Fulkerson JA, Eisenberg ME, Berge J. What's for dinner? Types of food served at family dinner differ across parent and family characteristics. Public Health Nutr 2012 Oct 19;17(1):145-155. [doi: 10.1017/s1368980012004594]
13. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
14. Resnicow K, Jackson A, Wang T, De AK, McCarty F, Dudley WN, et al. A motivational interviewing intervention to increase fruit and vegetable intake through Black churches: results of the Eat for Life trial. Am J Public Health 2001 Oct;91(10):1686-1693. [doi: 10.2105/ajph.91.10.1686] [Medline: 11574336]
15. Thorpe M. Motivational interviewing and dietary behavior change. J Am Diet Assoc 2003 Feb;103(2):150-151. [doi: 10.1016/s0002-8223(03)00005-1]
16. VanWormer JJ, Boucher JL. Motivational interviewing and diet modification: a review of the evidence. Diabetes Educ 2004 Sep 04;30(3):404-408. [doi: 10.1177/014572170403000309] [Medline: 15208841]
17. Ryan R, Deci E. Overview of self-determination theory: an organismic-dialectical perspective. In: Handbook of Self-determination Research. Rochester, NY, United States: University of Rochester Press; 2002.
18. Shingleton R, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
19. Miller ST, Marolen KN, Beech BM. Perceptions of physical activity and motivational interviewing among rural African-American women with type 2 diabetes. Womens Health Issues 2010;20(1):43-49 [FREE Full text] [doi: 10.1016/j.whi.2009.09.004] [Medline: 19944621]
20. Neumark-Sztainer D, Wall M, Perry C, Story M. Correlates of fruit and vegetable intake among adolescents. Findings from Project EAT. Prev Med 2003 Sep;37(3):198-208. [doi: 10.1016/s0091-7435(03)00114-2] [Medline: 12914825]
21. Larson N, Neumark-Sztainer D, Hannan P, Story M. Family meals during adolescence are associated with higher diet quality and healthful meal patterns during young adulthood. J Am Diet Assoc 2007 Sep;107(9):1502-1510. [doi: 10.1016/j.jada.2007.06.012] [Medline: 17761227]
22. Spruijt-Metz D. Etiology, treatment and prevention of obesity in childhood and adolescence: a decade in review. J Res Adolesc 2011 Mar;21(1):129-152 [FREE Full text] [doi: 10.1111/j.1532-7795.2010.00719.x] [Medline: 21625328]
23. Donelan K, Hill CA, Hoffman C, Scoles K, Feldman PH, Levine C, et al. Challenged to care: informal caregivers in a changing health system. Health Aff (Millwood) 2002 Jul;21(4):222-231. [doi: 10.1377/hlthaff.21.4.222] [Medline: 12117133]
24. Bird C. Gender differences in the social and economic burdens of parenting and psychological distress. J Marriage Fam 1997 Nov;59(4):809-823. [doi: 10.2307/353784]
25. Cousino M, Hazen R. Parenting stress among caregivers of children with chronic illness: a systematic review. J Pediatr Psychol 2013 Sep;38(8):809-828. [doi: 10.1093/jpepsy/jst049] [Medline: 23843630]
26. Barros L, Greffin K. Supporting health-related parenting: a scoping review of programs assisted by the Internet and related technologies. Estud psicol (Campinas) 2017 Sep;34(3):331-344. [doi: 10.1590/1982-02752017000300002]
27. Russell M, Harris B, Gockel A. Parenting in poverty: perspectives of high-risk parents. J Child poverty 2008 Mar 04;14(1):83-98. [doi: 10.1080/10796120701871322]

28. Storer H, Barkan S, Sherman E, Haggerty K, Mattos L. Promoting relationship building and connection: adapting an evidence-based parenting program for families involved in the child welfare system. Child Youth Serv Rev 2012 Sep 01;34(9) [FREE Full text] [doi: 10.1016/j.childyouth.2012.05.017] [Medline: 24347754]
29. Garcia A, DeNard C, Ohene S, Morones S, Connaughton C. “I am more than my past”: parents' attitudes and perceptions of the positive parenting program in child welfare. Child Youth Serv Rev 2018 May;88:286-297. [doi: 10.1016/j.childyouth.2018.03.023]
30. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. ACM Trans Comput Hum Interact 2020 Aug 31;27(4):1-27. [doi: 10.1145/3373759]
31. Zubatiy T, Vickers K, Mathur N, Mynatt E. Empowering dyads of older adults with mild cognitive impairment and their care partners using conversational agents. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan URL: https://dl.acm.org/doi/proceedings/10.1145/3411764 [doi: 10.1145/3411764.3445124]
32. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
33. Lester J, Branting K, Mott B. Conversational agents. In: The Practical Handbook of Internet Computing. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2004.
34. Olafsson S, O'Leary T, Bickmore T. Coerced change-talk with conversational agents promotes confidence in behavior change. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 Presented at: PervasiveHealth'19: The 13th International Conference on Pervasive Computing Technologies for Healthcare; May 20 - 23, 2019; Trento Italy. [doi: 10.1145/3329189.3329202]
35. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2:197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
36. Beneteau E, Guan Y, Richards OK, Zhang MR, Kientz JA, Yip J, et al. Assumptions checked. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-23. [doi: 10.1145/3380993]
37. Beneteau E, Kientz J, Richards O, Yip J, Zhang M, Hiniker A. Communication breakdowns between families and Alexa. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300473]
38. Ammari T, Kaye J, Tsai JY, Bentley F. Music, search, and IoT. ACM Trans Comput Hum Interact 2019 Jun 06;26(3):1-28. [doi: 10.1145/3311956]
39. Shin J, Huh-Yoo J. Designing everyday conversational agents for managing health and wellness: a study of Alexa skills reviews. In: Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2020 Presented at: PervasiveHealth '20: 14th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 18 - 20, 2020; Atlanta GA USA.
40. Bentley F, Luvogt C, Silverman M, Wirasinghe R, White B, Lottridge D. Understanding the long-term use of smart speaker assistants. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-24. [doi: 10.1145/3264901]
41. Sciuto A, Saini A, Forlizzi J, Hong J. "Hey Alexa, what's up?": a mixed-methods studies of in-home conversational agent usage. In: Proceedings of the 2018 Designing Interactive Systems Conference. 2018 Presented at: DIS '18: Designing Interactive Systems Conference 2018; Jun 9 - 13, 2018; Hong Kong China.
42. Yarosh S, Thompson S, Watson K, Chase A, Senthilkumar A, Yuan Y, et al. Children asking questions: speech interface reformulations and personification preferences. In: Proceedings of the 17th ACM Conference on Interaction Design and Children. 2018 Presented at: IDC '18: Interaction Design and Children; Jun 19 - 22, 2018; Trondheim Norway. [doi: 10.1145/3202185.3202207]
43. Lovato S, Piper A, Wartella E. Hey Google, do unicorns exist?: conversational agents as a path to answers to children's questions. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children. 2019 Presented at: IDC '19: Interaction Design and Children; Jun 12 - 15, 2019; Boise ID USA. [doi: 10.1145/3311927.3323150]
44. Jo E, Bang H, Ryu M, Sung EJ, Leem S, Hong H. MAMAS: supporting parent--child mealtime interactions using automated tracking and speech recognition. Proc ACM Human Comput Interact 2020 May 28;4(CSCW1):1-32. [doi: 10.1145/3392876]
45. Musher-Eizenman D, Holub S. Comprehensive feeding practices questionnaire: validation of a new measure of parental feeding practices. J Pediatr Psychol 2007 Sep 09;32(8):960-972. [doi: 10.1093/jpepsy/jsm037] [Medline: 17535817]
46. Savage JS, Fisher JO, Birch LL. Parental influence on eating behavior: conception to adolescence. J Law Med Ethics 2021 Jan 01;35(1):22-34. [doi: 10.1111/j.1748-720x.2007.00111.x]
47. Buckley M, Cowan C, McCarthy M. The convenience food market in Great Britain: convenience food lifestyle (CFL) segments. Appetite 2007 Nov;49(3):600-617. [doi: 10.1016/j.appet.2007.03.226] [Medline: 17537540]
48. Horwath CC, Schembre SM, Motl RW, Dishman RK, Nigg CR. Does the transtheoretical model of behavior change provide a useful basis for interventions to promote fruit and vegetable consumption? Am J Health Promot 2013;27(6):351-357 [FREE Full text] [doi: 10.4278/ajhp.110516-QUAN-201] [Medline: 23398135]

49. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Aug 26;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
50. Diclemente C, Velasquez M. Motivational interviewing and the stages of change. In: Motivational Interviewing. New York, United States: The Guilford Press; Jan 2002.
51. Prochaska J, Redding CA, Evers KE. The transtheoretical model and stages of change. In: Health Behavior: Theory, Research, and Practice. Hoboken, New Jersey, United States: Jossey-Bass/Wiley; 2015.
52. Cassin S, von Ranson KM, Heng K, Brar J, Wojtowicz A. Adapted motivational interviewing for women with binge eating disorder: a randomized controlled trial. Psychol Addict Behav 2008 Sep;22(3):417-425. [doi: 10.1037/0893-164X.22.3.417] [Medline: 18778135]
53. Treasure J, Schmidt U. Motivational interviewing in the management of eating disorders. In: Motivational Interviewing in the Treatment of Psychological Problems. New York: The Guilford Press; 2008.
54. Rhee K, McEachern R, Jelalian E. Parent readiness to change differs for overweight child dietary and physical activity behaviors. J Acad Nutr Diet 2014 Oct;114(10):1601-1610 [FREE Full text] [doi: 10.1016/j.jand.2014.04.029] [Medline: 24953789]
55. Taveras E, Mitchell K, Gortmaker S. Parental confidence in making overweight-related behavior changes. Pediatrics 2009 Jul;124(1):151-158. [doi: 10.1542/peds.2008-2892] [Medline: 19564295]
56. Bolling C, Crosby L, Boles R, Stark L. How pediatricians can improve diet and activity for overweight preschoolers: a qualitative study of parental attitudes. Acad Pediatr 2009;9(3):172-178 [FREE Full text] [doi: 10.1016/j.acap.2009.01.010] [Medline: 19450777]
57. Muscat A. Ready, set, go: the transtheoretical model of change and motivational interviewing for “fringe” clients. J Employ Couns 2005 Dec;42(4):179-191. [doi: 10.1002/j.2161-1920.2005.tb01089.x]
58. Dray J, Wade T. Is the transtheoretical model and motivational interviewing approach applicable to the treatment of eating disorders? A review. Clin Psychol Rev 2012 Aug;32(6):558-565. [doi: 10.1016/j.cpr.2012.06.005] [Medline: 22819997]
59. Wilson GT, Schlam T. The transtheoretical model and motivational interviewing in the treatment of eating and weight disorders. Clin Psychol Rev 2004 Jul;24(3):361-378. [doi: 10.1016/j.cpr.2004.03.003] [Medline: 15245836]
60. Choi M, Lee Y, Heo Y, Hyun T, Lyu E, Park H, et al. Association between the frequency of dining out and the risk of obesity, diabetes mellitus, and dyslipidemia among Korean adults. Ecol Food Nutr 2019 Jul 21;58(6):560-574. [doi: 10.1080/03670244.2019.1644327] [Medline: 31327256]
61. Darmon N, Drewnowski A. Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. Nutr Rev 2015 Oct 25;73(10):643-660 [FREE Full text] [doi: 10.1093/nutrit/nuv027] [Medline: 26307238]
62. Rosinger A, Herrick K, Gahche J, Park S. Sugar-sweetened beverage consumption among U.S. adults, 2011-2014. NCHS Data Brief 2017 Jan [FREE Full text]
63. Freitas A, Albuquerque G, Silva C, Oliveira A. Appetite-related eating behaviours: an overview of assessment methods, determinants and effects on children's weight. Ann Nutr Metab 2018 May 29;73(1):19-29 [FREE Full text] [doi: 10.1159/000489824] [Medline: 29843129]
64. Vainik U, García-García I, Dagher A. Uncontrolled eating: a unifying heritable trait linked with obesity, overeating, personality and the brain. Eur J Neurosci 2019 Aug 25;50(3):2430-2445. [doi: 10.1111/ejn.14352] [Medline: 30667547]
65. Engin A. Eat and death: chronic over-eating. In: Obesity and Lipotoxicity. Cham: Springer; 2017.
66. Trinh H, Shamekhi A, Kimani E, Bickmore T. Predicting user engagement in longitudinal interventions with virtual agents. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: IVA '18: International Conference on Intelligent Virtual Agents; Nov 5 - 8, 2018; Sydney NSW Australia. [doi: 10.1145/3267851.3267909]
67. Teach, Learn, and Make with Raspberry Pi. Raspberry Pi Foundation. URL: https://www.raspberrypi.org/ [accessed 2020-09-16]
68. Python homepage. Python. URL: https://www.python.org/ [accessed 2022-04-19]
69. Bray T. The javascript object notation (json) data interchange format. IETF Datatracker. 2020 Jan 21. URL: https://datatracker.ietf.org/doc/rfc7159/ [accessed 2022-04-19]
70. Dialogflow. Google Cloud. URL: https://cloud.google.com/dialogflow [accessed 2022-04-19]
71. What is Alexa? Amazon. URL: https://developer.amazon.com/en-GB/alexa [accessed 2022-04-19]
72. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-09-16]
73. Zoom: Video conferencing, web conferencing. Zoom. URL: https://zoom.us/ [accessed 2022-09-29]
74. Craigslist Kolkata. Craigslist. URL: https://craigslist.org/ [accessed 2020-09-16]
75. Lewis JR. The system usability scale: past, present, and future. Int J Human Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
76. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/obesity/basics/adult-defining.html#:~:text=Adult%20Body%20Mass%20Index&text=If%20your%20BMI%20is%20less,falls%20within%20the%20obesity%20range [accessed 2022-09-29]

77. Childhood obesity facts. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/obesity/data/childhood.html [accessed 2020-09-16]
78. Strauss A, Corbin J. Grounded theory methodology: an overview. In: Handbook of Qualitative Research. Thousand Oaks, California: SAGE Publications; 1994.
79. NVivo Qualitative Data Anaysis Software. NVivo. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-09-29]
80. Holtzblatt K, Beyer H. Contextual Design Defining Customer-Centered Systems. Amsterdam, Netherlands: Elsevier; 1997.
81. Jake-Schoffman DE, Turner-McGrievy G, Wilcox S, Moore JB, Hussey JR, Kaczynski AT. The mFIT (Motivating Families with Interactive Technology) study: a randomized pilot to promote physical activity and healthy eating through mobile technology. J technol behav sci 2018 Mar 21;3(3):179-189. [doi: 10.1007/s41347-018-0052-8]
82. Beneteau E, Boone A, Wu Y, Kientz J, Yip J, Hiniker A. Parenting with Alexa: exploring the introduction of smart speakers on family dynamics. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376344]
83. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
84. Porcheron M, Fischer J, Reeves S, Sharples S. Voice interfaces in everyday life. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada.
85. Yip J, Sobel K, Gao X, Hishikawa A, Lim A, Meng L, et al. Laughing is scary, but farting is cute: a conceptual model of children's perspectives of creepy technologies. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300303]
86. Garg R, Sengupta S. Conversational technologies for in-home learning: using co-design to understand children's and parents' perspectives. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376631]
87. Cabitza F, Simone C. On-line questionnaires as lightweight tools to extract collective medical knowledge from daily practice. In: Proceedings of the Workshop “Collective Intelligence In Organizations: Toward a Research Agenda". 2010 Presented at: “Collective Intelligence In Organizations: Toward a Research Agenda"; Feb 6, 2010; Savannah, Georgia, USA.
88. Crawford K, Calo R. There is a blind spot in AI research. Nature 2016 Oct 20;538(7625):311-313 [FREE Full text] [doi: 10.1038/538311a] [Medline: 27762391]
89. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection companion. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Jul 05;2(2):1-26. [doi: 10.1145/3214273]
90. Cho E, Sundar S, Abdullah S, Motalebi N. Will deleting history make Alexa more trustworthy?: effects of privacy and content customization on user experience of smart speakers. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376551]
91. Wang J, Yang H, Shao R, Abdullah S, Sundar S. Alexa as coach: leveraging smart speakers to build social agents. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376561]
92. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
93. Kim J, Kim W, Nam J, Song H. "I can feel your empathic voice": effects of nonverbal vocal cues in voice user interface. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3334480.3383075]
94. Ciechanowski L, Przegalinska A, Magnuski M, Gloor P. In the shades of the uncanny valley: an experimental study of human–chatbot interaction. Future Generation Comput Syst 2019 Mar;92:539-548 [FREE Full text] [doi: 10.1016/j.future.2018.01.055]
95. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
96. 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 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]

97. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
98. Fried Y, Slowik LH. Enriching goal-setting theory with time: an integrated approach. Acad Manag Rev 2004 Jul;29(3):404-422. [doi: 10.5465/amr.2004.13670973]
99. Clark L, Pantidi N, Cooney O, Doyle P, Garaialde D, Edwards J, et al. What makes a good conversation?: challenges in designing truly conversational agents. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300705]

Abbreviations

CA: conversational agent
MI: Motivational Interviewing
MICA: Motivational Interviewing Conversational Agent
TTM: transtheoretical model

Edited by A Kushniruk; submitted 21.04.22; peer-reviewed by D Chrimes, B Nievas Soriano; comments to author 18.07.22; revised version received 05.08.22; accepted 06.08.22; published 07.10.22.

Please cite as:
Smriti D, Kao TSA, Rathod R, Shin JY, Peng W, Williams J, Mujib MI, Colosimo M, Huh-Yoo J
Motivational Interviewing Conversational Agent for Parents as Proxies for Their Children in Healthy Eating: Development and User Testing
JMIR Hum Factors 2022;9(4):e38908
URL: https://humanfactors.jmir.org/2022/4/e38908
doi:10.2196/38908
PMID:36206036

Original Paper

Examining Use Behavior of a Goal-Supporting mHealth App in Primary Care Among Patients With Multiple Chronic Conditions: Qualitative Descriptive Study

Farah Tahsin[1], MHsc; Tujuanna Austin[1], MSc; Brian McKinstry[2], MD; Stewart W Mercer[3], BSc, MSc, MBChB, PhD; Mayura Loganathan[4], MD; Kednapa Thavorn[5,6,7], PhD; Ross Upshur[1,8], MD, PhD; Carolyn Steele Gray[1,8], MA, PhD

[1]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[2]Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom
[3]Advanced Care Research Centre, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom
[4]Mount Sinai Academic Family Health Team, Mount Sinai Hospital, Toronto, ON, Canada
[5]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[6]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
[7]Faculty of Pharmacy, Chiang Mai University, Chiang Mai, Thailand
[8]Bridgepoint Collaboratory for Research and Innovation, Lunenfeld-Tanenebaum Research Institute, Sinai Health, Toronto, ON, Canada

Corresponding Author:
Farah Tahsin, MHsc
Institute of Health Policy, Management, and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College St 4th Floor
Toronto, ON, M5T 3M6
Canada
Phone: 1 647 825 4684
Email: farah.tahsin@mail.utoronto.ca

Abstract

Background: Although mobile health (mHealth) apps are increasingly being used to support patients with multiple chronic conditions (multimorbidity), most mHealth apps experience low interaction and eventual abandonment. To tackle this engagement issue, when developing an mHealth program, it is important to understand the social-behavioral factors that affect patients' use behavior.

Objective: The aim of this study was to explore the social and behavioral factors contributing to patients' use behavior of an mHealth app called the electronic Patient-Reported Outcome (ePRO). The ePRO app supports goal-oriented care delivery in interdisciplinary primary care models.

Methods: A descriptive qualitative study was used to analyze interview data collected for a larger mixed methods pragmatic trial. The original 15-month trial was conducted in 6 primary care teams across Ontario, Canada, between 2018 and 2019. The eligibility criteria for patients were being aged ≥60 years with ≥10 visits within the previous 12 months of study enrollment. For this analysis, patients were classified as long-term or short-term users based on their length of use of the ePRO app during the trial. The Social Cognitive Theory by Bandura was used to categorize social-behavioral factors that contributed to patients' decision to continue or discontinue using the app.

Results: The patient-provider relationship emerged as a key factor that shaped patients' experiences with the app and subsequent decision to continue using the app. Other factors that contributed to patients' decision to continue using the app were personal and social circumstances, perceived usefulness, patients' previous experience with goal-related behaviors, and confidence in one's capability. There was an overlap of experience between long- and short-term app users but, in general, long-term users perceived the app to be more useful and their goals to be more meaningful than short-term app users. This observation was complicated by the fact that patient health-related goals were dynamic and changed over time.

Conclusions: Complex patients' use behavior of a goal-supporting mHealth app is shaped by an array of sociobehavioral factors that can evolve. To tackle this dynamism, there should be an emphasis on creating adaptable health technologies that are easily customizable by patients and able to respond to their changing contexts and needs.

Trial Registration: ClinicalTrials.gov NCT02917954; https://clinicaltrials.gov/ct2/show/NCT02917954

(JMIR Hum Factors 2022;9(4):e37684) doi:10.2196/37684

KEYWORDS

mobile health; mHealth; multimorbidity; chronic disease management; goal-oriented care; multimorbid; app; primary care; telemedicine; use; usability; human factors; behavior; sociobehavioral; health technology; mobile phone

Introduction

Background

Mobile health (mHealth) apps are being increasingly used to deliver care and support patients with chronic conditions [1-3]. Managing chronic conditions effectively is an ongoing task that often requires sustained support from an interdisciplinary team of health care providers. Continued involvement of multiple health care providers in supporting chronic disease management can be costly to the health system [4] and demands the time and resources of providers as well as their patients [5]. This management may be particularly challenging for patients with complex needs. Individuals with complex care needs are those who live with multiple chronic conditions (multimorbidity) [6] and experience additional health- and biopsychosocial-related challenges because of increased treatment requirements, reduced functional ability, and socioeconomic challenges [7]. To improve patient experience and tackle the issue of high health care burden and costs, mHealth is considered to be an effective and efficient solution [8]. mHealth offers an array of functionalities, which can include remote monitoring of patients' vital signs and symptoms; ongoing and timely communication with multiple providers; and information sharing such as scheduled appointments, drug prescriptions, and renewals [9].

The positive benefits of using apps and web-based platforms to support complex patients is documented in the current literature [10,11]. For example, when patients with chronic illnesses use mHealth apps, they are more likely to engage in health-promoting behaviors such as fruit consumption and physical activity [12]. However, the benefits are more likely to be realized if the technologies are used as intended. Instead, most mHealth apps experience high attrition [13,14], defined as when an individual disengages from a technology-based intervention after initially committing to using the technology [15].

Objectives

Attrition has been considered a major challenge in mHealth-based interventions [15-17]. Previous research has identified that only a small number of participants use mHealth apps in the long term, and most patients abandon the app after a short period [16,18]. The reasons behind high attrition vary. Critical factors that drive attrition can include available social support and capital, trust in technology, intention, and ability to use the app [19-21]. A meta-analysis evaluating the rates of attrition in mHealth interventions showed that many attrition-focused studies had relatively short intervention (<2 months) or follow-up periods [15]. This synthesis work suggests a need to advance knowledge of attrition by exploring sociocognitive factors that contribute to patients' app use in the long-term and real-world settings [15]. To address this gap in the literature, this study explores community-dwelling patients' perception of the long-term use of an mHealth app by applying the Social Cognitive Theory (SCT) by Bandura [22] to unpack sociocognitive factors that play a role. The research question informing this study is as follows: *What are the social and behavioral factors that contribute to continued or discontinued use of a goal-management app tailored for patients with complex chronic conditions?*

Methods

Description of the Electronic Patient-Reported Outcome Intervention

The electronic Patient-Reported Outcome (ePRO) tool is both an mHealth app and a portal that enables goal-oriented care delivery by facilitating goal creation and monitoring by patients with complex chronic conditions working in collaboration with an interdisciplinary primary care team [9,23,24]. User-centered co-design methods were used to develop the app through multiple iterations [25]. The co-design method was operationalized using inputs from patients with complex care needs, caregivers, and the primary care team [25,26]. The usability and feasibility of the app were assessed during usability testing [25] and exploratory trials [26] of ePRO. The findings from the exploratory trial informed the modification of the ePRO app to meet patients' needs.

In a usability study of the ePRO trial, it was found that the app experienced gradual attrition of participants despite the tool scoring moderate usability [27]. The qualitative analysis presented in this paper was conducted to deeply explore social-behavioral factors that may be influencing patients' low engagement with the ePRO app that were found in the usability study. Of note, ePRO is not an open-source app and was only available to study participants for the duration of the trial. A screenshot of the app interface can be found in a previous publication [23].

Description of the Study Design

We conducted a descriptive qualitative substudy drawing on patients' interview data collected as part of a larger 15-month multisite pragmatic stepped-wedge trial of the ePRO tool [23,24]. The trial registration number for the study is NCT02917954. Following the stepped-wedge trial design, 6

sites were randomized into 2 intervention clusters, and 2 different clusters received the ePRO intervention at 2 different time points. As a result, the first group used the ePRO app for 12 months after a 3-month control period, and the second group used the app for 9 months after a 6-month control period [23]. The qualitative descriptive approach seeks to present data as close to how the participant would understand the phenomenon as possible, referred to as "*staying close to the data*." [28] This approach allowed us to present the patient's direct description of their experience of the ePRO intervention and the factors that they perceive as contributing to their discontinuation or continuation of use without too many interpretive interferences from researchers [28]. Therefore, the findings of this study closely represent patients' experiences with the intervention.

A 2-stage sampling strategy was used to recruit participants for the study. First, we recruited family health teams (FHTs), and then we recruited patients within each FHT. FHTs are designed to provide integrated, multidisciplinary primary care and are typically led by physicians or nurse practitioners [29]. A purposeful sampling strategy [30] was used to recruit 6 FHTs across geographically diverse areas (urban, rural, and suburban) of Ontario, Canada, from 2018 to 2019; this FHT recruitment process is described in detail in another publication [23,30]. The categorization of sites into rural, urban, and suburban settings was consistent with Statistics Canada's definition of rurality [31]. The geographic location of the FHT was important for capturing the variation in the study participants. The eligibility criteria for FHTs were being an Ontario-based FHT and willingness to participate in the ePRO study. Ontario is the largest province in Canada, with the highest population density, and most services provided by primary care teams are funded by the Ministry of Health.

Quantitative data (surveys and chart audits) were collected from all 6 sites, whereas qualitative data were collected from 3 case sites [23,24]. At first, 67% (4/6) of the sites agreed to participate as case sites. However, 17% (1/6) of the sites dropped out because of low patient recruitment [23,24]. Patient interviews, demographic surveys, and research memos collected in these 3 case sites were used to answer the research question of this study.

Participants and Interviews

The eligibility criteria for the recruited patients within the FHTs were being aged ≥60 years with ≥10 visits to the FHT within the previous 12 months. A total of ≥10 visits [32] and an age of ≥60 years [26] were chosen as both factors are considered as an indicator of complexity of this study population and were used as a recruitment strategy for the exploratory trial [26].

Using FHT electronic medical records, eligible patients were identified. The list of eligible patients was then given to FHT providers to assess whether the patients met the following additional criteria: (1) perceived willingness to engage in a conversation about goals of care, (2) ability to use a smartphone or tablet in English or having a caregiver who could do this on their behalf, (3) capability to provide consent to participate, and (4) willingness to complete surveys every 3 months thereafter until the trial concluded. Eligible patients were approached by their FHT staff (ie, care coordinators and administrators) and asked if they would be willing to speak to a research team member about the project. Recruitment occurred during a scheduled office visit or by phone. A detailed description of the recruitment procedure has been provided elsewhere [23,24].

Patients' demographic information was collected through a survey at the beginning of the study. The first set of interviews was conducted at the midpoint of the trial, 4 to 6 months after the patients started using the app (the timing of the interviews depended on whether they were in the 12- or 9-month use group). The second round of interviews was conducted at the end of the trial. The purpose of the 2 sets of semistructured interviews was to explore patients' overall experience with the ePRO intervention and how that experience changed over time. The semistructured interview guide addressed the following topics: (1) perception and experience of using the ePRO app, (2) patients' relationship with their care team, (3) perception and experience of setting goals through ePRO, and (4) impact of ePRO on patients' daily lives. Following the first set of interviews, the semistructured interview guide of the study was modified for the second set of interviews. Findings from the first set of interviews guided the iteration process for the semistructured interview guide and were decided by the research team members (FT [research coordinator], TA [graduate research assistant], JS [research coordinator], and CSG [research scientist and principal investigator with extensive qualitative research experience]).

The interviews were 25 to 40 minutes long and were conducted by 1 of 4 research team members (FT, TA, JS, and CSG). Each interview was audiotaped and transcribed using a commercial transcription service. Transcripts were checked for accuracy against recordings by a member of the research team.

Ethics Approval

Ethics approval was received from the University of Toronto Health Sciences Research Ethics Board (33944) and the research ethics boards of the 3 participating primary care practices. All patient participants provided informed verbal and written consent before initiation of study activities.

The Theoretical Framework for Data Analysis

Multiple theories and frameworks have been used to explore the relationship between patients' social-behavioral factors and mHealth or eHealth use [13,32]. One such theory is the SCT by Bandura [33], which explains human behaviors through a model of interactions among behavioral, environmental, and social factors. This model has been used extensively to uncover which social and behavioral constructs may influence patients' use behavior of an mHealth app [33-35]. Textbox 1 shows the 5 key domains of the SCT. SCT is particularly well suited to examine patients' use behavior of an mHealth app such as ePRO as this app enables users to evaluate and monitor their goals over time and modify their behavior [23]. SCT also allowed us to understand the social-cognitive–related factors that contribute to the process by which patients decide to continue or discontinue app use. For example, one of the SCT domains, reciprocal determinism, is helpful to identify how personal, environmental, and behavioral factors can influence one's decision to continue or discontinue app use. Similarly, the

behavior capability and goal efficacy domains were helpful to identify how one's skills and confidence can influence their decision on app use.

Textbox 1. Domains of Social Cognitive Theory (SCT) [22].

Reciprocal determinism

- This constitutes the dynamic and reciprocal interaction of person (individual with a set of learned experiences), environment (external social context, technology, and aids), and behavior (responses to stimuli to achieve goals). In SCT, these components—behavior, environment, and individual—are seen as acting bidirectionally.

Behavior capability

- This constitutes a person's actual ability to perform a behavior through essential knowledge and skills.

Goal efficacy

- This constitutes the level of a person's confidence in their ability to successfully perform a behavior.

Use reinforcement

- The internal and external responses to a person's behavior affect the likelihood of continuing or discontinuing the behavior.

Outcome expectancies

- This constitutes the anticipated consequences of a person's behavior. Outcome expectancies can be health-related or not health-related.

This theory was used to guide data analysis to explore how complex patients' personal beliefs and attitudes and physical and social environmental factors affected their engagement pattern (long-term and short-term app use) with ePRO. Although SCT can be used as an explanatory framework, it was applied in this study to help categorize factors influencing use and relate those to engagement patterns. During the interview debrief sessions, memoing activities, and initial reading of the transcripts, the authors (TA, FT, and CSG) agreed that SCT demonstrated a fit with the interview data. As we chose SCT as the right analytical tool based on emerging interview data, we did not encounter the challenge of forcing data into categories.

A combination of 2 techniques was used to analyze the study data. In stage 1, the transcripts were inductively coded by 2 analysts (FT and TA). During the analysis, the research team met to discuss the identified codes and resolve any coding discrepancies. After coding 4 transcripts, the team decided that the coding scheme was appropriate. We reached data saturation after coding 12 transcripts. Data saturation was determined when no new codes emerged from the transcripts [36]. After coding all 22 transcripts, the codes were mapped onto the SCT categories, meaning that inductively identified codes were plotted within the SCT categories to form themes [37].

The first stage allowed us to see the social and behavioral factors related to use. However, to see how these factors related to each other and changed over time, we engaged in the second analysis stage of restorying.

Restorying is defined as the method of rewriting participants' oral data temporally to draw a link between previous experience and subsequent experiences [38]. Restorying revealed how themes related to each other and changed over time. It also allowed us to more clearly see pattern differences across different user groups (short- vs long-term users), which allowed us to more directly address our question regarding social and behavioral factors that were related to continued or discontinued use. Restorying allowed us to generate exemplary narratives of long- and short-term app users as a means to illustrate these patterns. The definitions of long-term and short-term app users are described in the following section.

To restory patient data, 2 analysts (FT and TA) constructed a matrix of themes that distinguished between long-term and short-term app users (Table S1 in Multimedia Appendix 1 [38-40]). After examining both columns of long- and short-term app users, 2 research team members created a storyline for each group that captured the experience of the overall group. The restorying allowed us to see the connections between SCT constructs within the context of patient use of ePRO and how those connections influenced use progression over 15 months [38]. Although one of the major criticisms of SCT is that it does not recognize the wider social structure that influences an individual's use behavior [39], the analytic method of restorying addresses this challenge by highlighting the social contexts influencing use behavior over time. A detailed description of the 2-stage method can be found in Multimedia Appendix 1.

To enhance the rigor of this study, the researchers undertook several strategies to increase the credibility and trustworthiness of the findings [41]. The research team members met regularly to discuss codes and findings. In addition, throughout the restorying process, both researchers discussed the accuracy of the storyline. Member checking [36] was conducted with study participants to examine the accuracy of the 2 storylines and the overall interpretation of the study findings. Furthermore, having 2 data analysts helped ensure the dependability of the findings [40]. Both analysts (FT and TA) had graduate-level training in qualitative data analysis. In addition, one team member (CSG) provided supervisory support during the analysis.

Categorizing Patients Into Long-term and Short-term App Users

On the basis of patients' app-automated use logs, patients were classified into 2 categories: long-term users and short-term users. Of the 22 interviewed individuals, 9 (41%) were short-term users and 13 (59%) were long-term users. Participants who did not use the ePRO app after initial onboarding or used it for <3 months were categorized into the "short-term user" group. By contrast, the participants who used the ePRO app for >3 months were categorized into the "long-term user" group. The 3-month cutoff period was determined because the app experienced a sharp decline in use at 3 months [27]. This 3-month cutoff period is also consistent with the previous literature [41].

Results

Overview

There were 44 study participants in the larger pragmatic trial, with 37 (84%) from the 3 case sites. Of the 37 patients who were invited to participate in the interviews, in total, 22 (59%) were interviewed. Of the 22 interviewed patients, 17 (77%) participated in both interviews, 3 (14%) participated in only the midpoint interviews, and 2 (9%) participated in the last interview only. A total of 41% (15/37) of the participants did not take part in the interviews because of scheduling issues, illness, being out of the country when the interview was scheduled, or not responding to interview requests.

Demographic Description of the Participants

The demographic information of the study participants can be found in Table 1. The mean age of the 22 interviewed participants was 75.1 (SD 5.67) years, and 45% (10/22) self-identified as female. We also reported the demographic characteristics of the participants who did not take part in the interviews (15/37, 41%) to show any demographic differences between the interviewed and noninterviewed groups. It is worth noting that there were more noninterviewed participants in the lowest income quintile. However, we did not identify any statistically significant demographic differences between the interviewed and noninterviewed participants. We conducted descriptive statistical analyses (2-tailed Student *t* test for continuous variables and Mann-Whitney *U* test for categorical variables) to explore the differences between the groups (interviewed and noninterviewed and short-term and long-term).

Table 1. Demographic characteristics (N=37).

Variable	Interviewed participants (n=22)	Noninterviewed participants (n=15)
Age (years), mean (SD)	75.1 (5.6)	71.14 (6.5)
Sex—female, n (%)	10 (45)	5 (33)
Smartphone comfort level score, mean (SD)[a]	2.17 (1.4)	3.64 (1.4)
Number of chronic conditions, mean (SD)	4.88 (2.1)	3.07 (1.8)
Family income, n (%)		
CAD $0 to $29,000 (US $0-$21,310.40)	1 (5)	4 (27)
CAD $30,000 to $59,000 (US $22,045.30-$43,355.70)	7 (32)	4 (27)
CAD $60,000 to $89,000 (US $44,090.60-$65,401)	3 (14)	3 (20)
>CAD $90,000 (US $66,135.90)	4 (18)	3 (20)
Education, n (%)		
Lower than high school	2 (9)	2 (13)
High school	2 (9)	4 (27)
Some college or university	4 (18)	3 (20)
University (undergraduate or graduate)	4 (18)	5 (33)

[a]The range of the smartphone comfort level score is 1 to 5. A higher score indicates a higher comfort level with a smartphone.

Summary Description of the Themes

The patient interviews revealed insights into the factors that influenced patients' decision to continue or discontinue app use. When discussing their use of the ePRO app, patients identified what encouraged them to use the app, including factors relating to their social and clinical relationships, capability to use the app and perform goal-related activities, and their expected outcomes from the ePRO app. Table 2 summarizes these factors in relation to the SCT domains. In addition, to provide a contextual understanding of these factors, long-term and short-term user narratives generated by restorying the data are first presented (Textbox 2), followed by a more in-depth exploration of each factor as it emerged in the full data set.

Table 2. Description of the themes.

Category and subcategory	Exemplary quotes	
	Long-term user	Short-term user
Reciprocal determinism[a]	"They [care team] always know what to do with me, so there was no problem setting goals because they know that I am trying to be active and healthy. and I kept using it (ePRO app) daily because I know they (care team) are watching my data." [Female, patient 18]	"I just did not know if anyone is looking at my data, there was no communication from you guys [research team] or my nurse or doctor here. There was no feedback for me about my data, so I felt like I am talking to the void when I was putting my information in. I would like to know if I was doing well or not. It would be helpful to talk to others (peers) about our goals, to see who else is doing the same thing as me and how they are feeling." [Female, patient 16]
Goal efficacy, behavior capability, and outcome expectancies[b]		
Subtheme 2a—confidence and skills in goals	"When my dietician first asked what goal I wanted to set, I knew it would be tracking my everyday walk, I knew it would be easy to keep up at because I have been doing this for long time. But ePRO made me more accountable, I wanted that accountability. I liked how the device asked me if I have achieved my goal for that day. Clicking yes to that felt good and I kept doing that." [Male, patient 7]	"Setting any goal was hard for me because my conditions flare up here and there and throws me off my routine. So I wasn't sure how well I can keep up with the goals...I sprained my ankle in last winter so then I was off my walking for 5 weeks. Considering all these troubles, I didn't work on my goals, and the app became redundant because what would I track. When the app asked Did I achieve my goal for the day, I did not want to keep saying no." [Male, patient 2]
Subtheme 2b—confidence and skills in technology	"I expected the app to have some direction for me about how I was doing on my goals, it was nice to see what I was accomplishing weekly basis. No complaints about the app, very easy to use...nothing complicated that anyone will have difficulty with...But I have used computer all my life for work so using this phone or any other phone is not a problem." [Female, patient 3]	"The small fonts or buttons in this phone [ePRO] was trouble...but I thought I will get used to it (the phone) but did not at the end. I was sometimes working on my goals but could not record it on the phone, so I lost interest in the phone...then I forgot about my goals too because I was not tracking it or doing anything about it." [Male, patient 21]
Subtheme 2c—outcome expectancies	"The main reason I enrolled-I wanted to stay on track of my goals and feel healthier over time-I thought the app was helpful to keep me on track." [Male, patient 1]	"When my doctor suggested this app, I did not know what to expect because there is nothing important, I need to work on, in my opinion anyway. My doctor suggested some goals but nothing very important...I could not make a purpose of it (ePRO)." [Male, patient 11]
Use reinforcement[c]	"I was bedridden so [provider's name] she was 'gung-ho' that I join her walking group for my recovery. And she said, "why don't you try this new thing we are doing, this will be good for you?". And She was right, it was nice to have the app because I know every Monday, I will have to say how many times I walked last week, so I tried to go out over weekends...She was there for me throughout, walking alongside me in every walking group." [Female, patient 6]	"My doctor did not think ePRO was helping me that much, because both of us thought I am doing fine without it, everything [diabetic symptoms] was on track, so we decided maybe I do not need it." [Male, patient 17]

[a]This domain refers to the dynamic relationship between individual, context, and behavior.

[b]This domain refers to individuals' confidence and skills in achieving their goals in the electronic Patient-Reported Outcome app and the perceived usefulness of the app.

[c]This domain refers to the internal or external responses that encourage or discourage behavior change.

Textbox 2. Long-term and short-term user stories.

Elaine: a long-term user story

Elaine considers herself to be a healthy individual whose diabetes symptoms are well managed through diet and exercise. She thinks of herself as "lucky" to have great health care providers who have helped her manage her symptoms for the past 2 and a half years. She has multiple other chronic conditions such as chronic pain and hypertension, but controlling diabetes symptoms is her foremost priority as she heard it can affect her other conditions. At first, she joined the electronic Patient-Reported Outcome (ePRO) study because her dietician at the family health team encouraged her to do so (*Social Cognitive Theory [SCT] domain: reciprocal determinism [social support]*). After talking to her dietician and talking to the ePRO study recruiter, Elaine agreed that ePRO would be a good addition to be more accountable toward her health-related goals (*SCT domain: outcome expectancies*). With her dietician, she decided on 3 goals that she always thought would be important to lead a healthy lifestyle. Elaine's goals were (1) lowering daily sugar intake, (2) joining walking programs with her peers facilitated by her dietitian, and (3) swimming every weekend in the local community center. She felt confident that she would be successful in achieving these goals as she had always been self-disciplined ("No TV from 9 AM to 6 PM") and kept a personal calendar to track her physical activity level. She also considers herself not in frail health, so she did not think that working toward those exercise-related goals would be hard for her (*SCT domain: goal efficacy and outcome expectancies*). She had also been working on those goals before the ePRO intervention, so she was confident that she had the necessary skills to work toward her goal (*SCT domain: behavior capability*) and thought ePRO would be beneficial for her to track those goals (*SCT domain: outcome expectancies*).

Elaine considers herself technologically savvy. However, she experienced a few technological difficulties while using ePRO. The most challenging one was being logged out of ePRO after taking a break from the tool during Christmas time when she visited her family in Scotland for 15 days. After not using ePRO while she was away, Elaine was locked out of the app. After returning from her holidays, she contacted her dietician to resolve the issue (*SCT domain: reciprocal determinism*). Her dietician asked her to contact the research team as she could not fix the technical issue for Elaine. Elaine's technical issue was resolved in 2 days, and she continued to use the app until the end of the study. In the final reflection, Elaine believed the app was good for her to be accountable toward her goal, and she derived satisfaction from that accountability. At the end of the study, Elaine planned to continue to track her goals through her calendar, which was how she tracked her goals before using the ePRO app. She thinks ePRO would benefit from having a communication feature. That way, she could communicate with her peers who are also using ePRO and working toward similar goals.

Josh: a short-term user story

Josh is a man aged 76 years with several concurrent chronic conditions, including diabetes, hypertension, and arthritis. Josh considers himself to have a fair understanding of his ailments and considers that his conditions are fairly well managed. Josh is the primary caregiver to his wife, who is ill. As a result of this caregiving role, Josh finds that he does not often have time to participate in social groups such as walking groups offered through his local community center (*SCT domain: reciprocal determinism [social support]*). Josh is a patient at a family health team where he has access to both primary care and allied health services. At the suggestion of his family physician, Josh agreed to participate in the ePRO study (*SCT domain: reciprocal determinism*). However, he did not expect the app to be useful as he considered himself to be "tech illiterate," so he did not think he would be able to use the app without his wife's help, and he did not think he had any important goals to work toward as he already had a healthy lifestyle (*SCT domain: behavior capability and outcome expectancies*).

In addition, Josh was hesitant to set a goal as he had never had a health-related goal before and was uncertain about whether he had the necessary skills or discipline to keep up with a specific goal (*SCT domain: behavior capability*), so he was not sure if ePRO would add value to his life (*SCT domain: outcome expectancies*). However, with assistance from the ePRO study team and his family physician, Josh created the following SMART goals: (1) eat at least one fruit every day and (2) walk for at least 10 minutes every day.

At the beginning of the study, Josh completed his check-in questions regularly. Over time, Josh began checking in on the app less and less, eventually not using the technology at all. When the ePRO study team reached out to Josh, he stated that he forgot his password and was unable to log in to the ePRO platform, so he did not use it. Although Josh describes himself as "computer illiterate," he found the ePRO app and web platform easy to use. Josh also found that, whenever he met with his health care provider, they did not discuss his goals but rather spoke about his medications and management of his conditions, resulting in goal setting becoming less of a priority (*SCT domain: behavior enforcement*). Josh found ePRO to be good for self-monitoring, but he did not find the technology useful for communicating with his health care team.

A major decision that was made during the analysis was to collapse 3 SCT domains—goal efficacy, behavior capability, and outcome expectancies—into one as it was identified that patients' confidence in their goal and technological skills was linked to the anticipated outcome of the ePRO app. Previous studies on goal-setting behavior have also identified that, in a real-world setting, individuals' confidence in health-related goals is confounded by their outcome expectancies, capability and skill level for carrying out various goals and activities, and technological and health literacy [39,42]. Applications of SCT in the literature have found that the relationship between multiple domains of SCT is multidirectional rather than unidirectional, as suggested by the original SCT, meaning that SCT domains can be both antecedents and consequences of each other [39]. For example, individuals who receive no feedback on their performance may lose motivation to continue engaging in a task and anticipate negative outcomes from their performance. Hence, in this analysis, we grouped these 3 domains together to retain the interrelationship as factors that contributed to patients' use behavior: confidence and skills in goals, confidence and skills in technology, and outcome expectancies.

Description of the Themes

Theme Overview

In this section, we elaborate on the themes identified in the data according to the SCT domains. Some domains had richer information than others. For example, the themes related to reciprocal determinism, goal efficacy, and outcome expectancies had more nuanced data compared with the other 2 themes, which were behavior capability and use encouragement.

Theme 1: Reciprocal Determinism

Reciprocal determinism focuses on the dynamic interaction between person-context behavior and the influence of this dynamic interaction on individuals' behavior. As demonstrated

in the long- and short-term user narratives as well as in Table 2, social and clinical relationships are key factors for the continued use of the ePRO app.

In total, 46% (6/13) of long-term users described their longstanding relationship with their primary care providers as being beneficial to setting meaningful goals:

> *I got lucked out with my providers, they will always know exactly how to deal with me and keep me out of the hospital, which is my main goal. My doctor knows that my nurse and dietitian here (primary care team) know that, so it was easy to set those goals to keep my blood sugar low.* [Long-term user, male, patient 7]

Short-term users also described a good relationship with their providers. However, 44% (4/9) of short-term users described that their providers did not discuss the ePRO app during their clinic visits. Participants listed the following reasons for not discussing the ePRO app with their providers: clinicians' heavy workload, not having enough time during the visit, feeling that it was unnatural to discuss the app during a regular clinic visit, and feeling that their goals were personal work and did not fall under providers' responsibility. A participant described the following:

> *Dr. [physician's name] is great, but he is really busy, so I did not want to waste his time talking about my walking schedule. He needs to check my blood pressure level; I would not bring up how many times I walked last month. Feels irrelevant for him to know that.* [Short-term user, male, patient 2]

Another way the patient-provider relationship influenced app use was when patients faced any sort of technical error in using the app or had to modify their goals after the initial goal-setting process. Specifically, long-term users were more likely to reach out for support and tended to report more instances of connection with their providers regarding the ePRO app. Some of the common technical challenges were (1) being logged out of the app because of prolonged inactivity, (2) forgetting passwords, and (3) inability to modify goals based on patients' needs. In terms of modifying goals, ePRO did not allow patients to modify their own goals, so primary care providers had to modify the goals for them. Therefore, when patients needed to modify their goals, they were uncertain about how to do that:

> *After they (government) changed the number of blood glucose tests I can do per week, my goal had to be changed because I wanted to test my glucose level daily but after they changed it, now I only test twice a week, but I still it report it on the phone just not daily. And my nurse over here changed it (frequency of reporting) for me.* [Long-term user, male, patient 12]

When faced with these technical difficulties or needed modifications, patients either abandoned the ePRO app or reached out to their health care providers or research team to solve the issue. Most long-term users (7/13, 54%) chose the latter option:

> *I was locked out of the app when I was on vacation...after I got back, I contacted the dietician over here (care team), and she connected me to you guys. Everything got resolved within 2 days, I kept using it.* [Long-term user, male, patient 1]

Short-term users, by contrast, decided to abandon the app and did not reach out for support when they faced similar technical difficulties:

> *It would be good if I could change my goals in the app because walking 5 km is what I set out to do at the beginning. It was too ambitious of a goal in this bad winter. I never reached 5 km, so I never had anything to report on the app...I did not reach out to my nurse practitioner, I guess I forgot about it (ePRO) for a while, and then I asked you (research team) to take it away.* [Short-term user, female, patient 22]

Both long- and short-term users also reflected on the fact that their relationships with peers and their communities could influence their app use behavior. For example, a patient discussed that being able to communicate with their peers would be useful in understanding others' experiences with the ePRO app:

> *Sometimes I felt that the app does not give me enough feedback. There could be more photos, a thumbs up if I did well. I'm a unique person so when I found I felt that way I thought, well I wonder if anyone else is feeling that way. So, communicating with other people that are using it without divulging your specific things would be nice.* [Long-term user, female, patient 19]

Importantly, unexpected changes in these relational contexts also influenced patients' use behavior, for example, a sudden transition to a caregiving role, a move away from social ties, or a divorce:

> *After my marriage fell apart, I moved to this area with my partner and I have to keep going back to the city to meet my friends, which makes it harder for me to meet people here. I am currently in an anxiety support group here, but I went off track with my other goals. I check the app (ePRO) sometimes but not regularly because I have nothing to report on.* [Long-term user, female, patient 14]

Theme 2: Goal Efficacy, Behavior Capability, and Outcome Expectancies

Overview

Patients' confidence, skills, and anticipated outcomes from the app influenced their use behavior. Although presented as distinct domains in SCT, data from this study suggest that the domains of goal efficacy, behavior capability, and outcome expectancies are linked.

The restorying work reveals these connections, which are best represented in the long- and short-term user narratives in Textbox 2. However, some participants' accounts also show that individuals' confidence in themselves to achieve goals (perceived goal efficacy), skills necessary to use the app

(behavior capability), and commitment to engage with the app to achieve set goals (outcome expectancies) are intertwined and influence each other. These outcome expectancies were also related to app functionality. This collapsed theme consisted of the following subthemes: (1) patients' confidence and skills with goals and their impact on ePRO use (subtheme 2a—confidence and skills with goals), (2) patients' confidence and skills in using technologies and their impact on ePRO use (subtheme 2b—confidence and skills in technology), and (3) patients' expected outcome from the ePRO app and its impact on their use behavior (subtheme 2c—outcome expectancies).

Subtheme 2a: Confidence and Skills With Goals

This subtheme demonstrates patients' descriptions of how their confidence in their goals and their skills to achieve the goals influenced their ePRO use behavior. Previous goal-setting experience and familiarity with goal-related tasks influenced patients' confidence in achieving the goals set in the ePRO app. Patients who had been working on a goal for a long time were more confident in their skills to achieve a goal. A total of 38% (5/13) of long-term users had already been working on a number of health-related goals before enrolling in the study and had been tracking their progress using electronic or paper-based tools such as calendars, wearable technologies, and handwritten notes. For these participants, the ePRO app was an additional electronic way to track their goals. These participants demonstrated confidence that they had the necessary skills to set appropriate goals and achieve them with the use of ePRO and, because they had the confidence and skills, they also had better outcome expectancy from the ePRO app:

> *I did pretty well in terms of crushing all my goals...because I already had the same goals, I was already continuing with the exercise program. So, it (ePRO goals) was just a continuation. I just kept up with the same tasks, swimming, walking that I was doing before joining your study.* [Long-term user, female, patient 3]

By contrast, patients who did not have any previous goal-setting experience reflected on the fact that setting a meaningful goal was difficult for them. Consequently, their providers had to suggest some goals for them, but some patients found that those goals were not personally meaningful. In these cases, not having previous goal-setting experience negatively affected patients' ability to set meaningful goals, which in turn affected their use behavior:

> *I've never had health goals before, so could not come up with one when they (health provider) asked me what I want to put in here (ePRO app). I got some kidney conditions, so my doctor suggested I set daily goals of drinking eight glasses of water and tracking them. I did not think I need to track it; I remember it anyway. I don't need a phone to tell me I need to hydrate. I did not think the goal was anything important for me to track on a phone.* [Short-term user, male, patient 11]

In terms of individuals' confidence in achieving their goals, some long-term users (6/13, 46%) indicated that their traits, such as "will-power," "self-discipline," and "motivation," boosted their confidence that they would be able to reach their goals:

> *It [achieving health goals] has nothing to do with the phone [ePRO app]. It has everything to do with the person. You have to be determined that you are going to walk. And you're going to set your goal—you're going to walk a block and you're going to walk back. You have to have determination. You have to have the willpower to say, I'm going to do it and that's it. ePRO is not going to do it for you, but it was good to have to see my progress. I thought it (ePRO) was a neat way to see how I am doing.* [Long-term user, female, patient 6]

In addition, patients reflected on the fact that their confidence and skills in achieving a goal changed over time depending on their health. When patients felt that they were not able to achieve their goals because of health and life circumstances and they did not have "enough" to report on the app, they discontinued using it:

> *Initially, I set up my goal to go 3 miles walking every day. But after my surgeries and my accident, there was no way I could do it. I was barely getting out to walk my dogs. I was falling short every day and it made no sense for me to use the app, I just felt sad that it [ePRO] kept showing me I was not the go-getter anymore. I did not know how to pause it [ePRO].* [Short-term user, female, patient 15]

Subtheme 2b: Confidence and Skills With Technology

Not surprisingly, patients who did not think that they had the necessary technological skills to use the ePRO app discontinued their use.

Several patients (14/22, 64%) discussed that they were technologically savvy enough to be able to use the app:

> *I found the app to be user-friendly, very clean, nothing too difficult, but I am good with computers and all that stuff, a tech-junkie. I use computers, phones, iPad all the time.* [Long-term user, female, patient 19]

Some participants (4/22, 18%) stated that they needed help using the ePRO app as often the fonts were too small:

> *I never had to use the computer for my work so never learned it. Now I got muscular dystrophy, so the fonts were way too small for me, so I did not use the app at all. I used the app [ePRO] on my computer, but I am not very good at it. My wife must help me a lot. I cannot even send an email; she will just do it for me. I ended up not using it [ePRO on the computer] at all.* [Short-term user, male, patient 17]

Subtheme 2c: Outcome Expectancies

Patients described their anticipated outcomes from the ePRO app. Typically, for long-term users, ePRO seemed like a beneficial addition to their health. A long-term user described that, while enrolling in the study, they anticipated that ePRO would make them more accountable toward their goals:

> *I wanted to get off my oxygen tank, I do not want to lug this machine everywhere. So I need to drop some pounds...by walking, exercising...I thought this phone would show me how I am doing, am I doing it too much, am I getting any good.* [Long-term user, female, patient 20]

By contrast, 33% (3/9) of the patients who were short-term users described that they discontinued using the app as they did not think that the app was "well-developed" to be implemented in the real world. Therefore, they did not think that the app would be a beneficial addition to their lives. A short-term user described their dissatisfaction with the functionality of the app:

> *I think that's all [research on people taking control over their health] a great idea I just feel that the actual implementation isn't as far advanced as it needs to be for it to work effectively, at least for me. I use my fitbit anyways to count my steps which is far better because that watch automatically counts my steps. I could not see any use for it [ePRO app] to work on my goals. I did not see any benefit for my health from it.* [Short-term user, male, patient 10]

Theme 3: Use Reinforcement

The use reinforcement domain of SCT suggests that internal and external factors such as internal satisfaction or external rewards can encourage or discourage individuals' behavior change. In total, 38% (5/13) of long-term users reported that they felt a sense of accomplishment (ie, internal reward) when they were able to "check off" their goals in the ePRO app. The app had the following question—"did you achieve your goal yesterday?"—and patients had the option of reporting yes or no. Some patients (6/13, 46%) found this exercise rewarding:

> *Well, to be honest, the only thing it did was—I do it [check off the list], used to do it every Monday morning, and it focused me on not smoking. That was the motivation every Monday morning, you know.* [Long-term user, female, patient 20]

Some short-term users (2/9, 22%) identified that they had already used many other legacy devices such as calendars, notebooks, cell phones, and glucose monitoring devices. These participants found reporting the same measures in 2 different tools to be redundant, and they did not think of the ePRO app as an important addition to their health-related goals:

> *I am an old school paper-pencil, calendar on refrigerator person, so that helps me to visualize my progress every day. I see them every day before breakfast, so I know what I had to do that day. The phone [ePRO] just stayed on my night table.* [Short-term user, female, patient 22]

An unexpected external influence can be discouragement from providers. Among the 9 short-term users, 22% (2/9) of the participants reported receiving advice from their providers to discontinue the use of ePRO. The factors that contributed to providers' discouragement were patients' frail health, patients' anxiety with the app regarding not being able to reach their goals, and changed health-related priorities:

> *My breathing issue has gotten worse in winter so I was not working on my goals anymore...When I told her [health provider] that I am worried about not reaching my goal, I feel anxious that I am not reaching my goal, she said "just forget about it [ePRO] for now, let's get back you to feeling good first," so I thought okay one thing off my list. I felt better.* [Short-term user, female, patient 13]

Long-term and Short-term User Stories

The 2 narratives presented in Textbox 2 offer a composite understanding of long- and short-term users of the ePRO app, linking elements of the stories shared by different participants to SCT domains.

Discussion

Principal Findings

This study used descriptive qualitative methods and restorying analytic techniques to explore the social and behavioral factors contributing to patients' use behavior of the ePRO tool. Study findings show that patient-provider relationships, patients' social relationships, and patients' personal circumstances play a central role in their decision to continue or discontinue the use of the ePRO app.

Leveraging SCT as a tool for data analysis, we were able to identify social-behavioral factors that contribute to patients' decision to continue or discontinue app use, such as their social and environmental factors and relationships (domain 1); confidence and skills in using technology, confidence and skills in setting and achieving goals, and expected outcomes from the intervention (domain 2); and encouraging factors (domain 3). Study data reveal that the SCT constructs of goal efficacy and behavior capability are also importantly related as capability and skill influence perceived confidence in completing a task. This interrelationship makes sense theoretically. SCT suggests that performing a behavior successfully increases individuals' confidence in their ability to accomplish goals as they believe that they have the skills to achieve goals through behavior change [22]. In addition, performing a behavior successfully also affects one's outcome expectancies as one believes that they have the skills and confidence to receive benefits from an action [22,43].

The stories show the themes of the interactions and links between concepts that the descriptive analysis could not. For example, an important interpretive theme that emerges from Josh and Elaine's stories is that patients' confidence and previous experience in goal setting influenced their capability and expectations from this goal-oriented intervention. Josh and Elaine approached their goals with varying degrees of experience, confidence, and attachment. For example, Elaine's previous experience with goal setting helped her feel more competent and skilled in achieving future goals, which subsequently increased her intention to track goals through ePRO, whereas Josh's lack of experience with goal setting made it challenging for him to make meaning of his goal, which translated into his reduced interest in tracking goals through ePRO.

Furthermore, the stories also show, in an interpretive manner, an important divergence in how long- and short-term users react to technical errors. App-related technical errors are ubiquitous, and many app-based interventions experience significantly high attrition after users experience an error [44]. As such, it is important to explore patients' strategies to mitigate risk and what factors contribute to their motivation to resolve such technical errors [24,45,46]. The patient-provider relationship emerged as an important mitigating factor when resolving technical errors. In Elaine's story, her strong relationship with her providers, the meaningfulness of her goals, and the satisfaction obtained from achieving goals influenced her motivation to proactively troubleshoot the problem and return to the app. This was a common occurrence among many long-term users, who would more readily troubleshoot technical errors with their primary care providers. Although this study provides an initial indication of the influence of the patient-provider relationship on technology use behavior, future studies should be conducted to determine the strength of this influence [47]. By contrast, for Josh, the combination of technical error and lack of meaning of his goals contributed to discontinuing app use. This finding shows that participants' goal-setting success was related to user experience with the app. If participants face difficulties using the app interface, they may abandon the goal-tracking exercise altogether, as demonstrated in Josh's story. In summary, factors such as the patient-provider relationship and app user experience can play an important role in a patient's decision to continue or discontinue using a goal-oriented app.

Another important study finding that emerged from the interview data is the importance of meaningful goal setting for an effective behavior change intervention. Hence, when setting patients' goals, a strong focus on patients' perception of the meaningfulness or fit of the goal in their daily lives should be accounted for as this meaningfulness of the goal can influence not only behavior change but also patients' adherence to a newly adopted technology [48,49]. This goal-oriented conversation between the patient and provider should also include an exploration of goal-setting and goal-monitoring tools that the patient may already be using, such as calendars, health-monitoring devices, or personal phones, as the study data suggest that often patients prefer devices and tools that they are familiar with rather than adopting a new tool [50].

Comparison of Themes With Previous Research

The findings of this study support previous study findings that health technologies are often discontinued and abandoned when they lack features of meaningful customization and is not part of users' already existing devices such as personal phones [50]. In addition, the study findings suggest that health-related goals change over time for patients with multiple chronic conditions, so designing apps that offer patient-driven customization and modification techniques will be helpful in repurposing the same technology at multiple time points of the life cycle. For example, patient 15 shared that their ability to achieve their goals changed over time because of emerging health issues, but they were unsure of how to modify the goals in the ePRO app. This design feature in ePRO was intentional based on a previous exploratory trial of the app (which was <4 months) [26]. In the exploratory trial of ePRO, it was found that the patients preferred provider consultations while changing their goals; hence, the app required the providers to modify goals on behalf of the patients. However, in this longer pragmatic trial of ePRO in which patients used ePRO for 9 to 12 months, patients preferred to modify their goals on their own, as demonstrated in these study findings. This contradiction may be due to the prolonged use of ePRO; for example, with prolonged use, patients' confidence in using the app changed, which in turn helped them feel like they could take charge of their goals. This finding demonstrates the importance of longitudinal evaluation of mHealth apps compared with a shorter follow-up time as patients' confidence, skills, and health needs from the app change over time, which may not be captured in a shorter trial [15].

For example, previous studies with shorter follow-up periods have identified that factors such as health literacy, motivation, capabilities, social and environmental structures, and social support have an impact on mHealth engagement [51,52]. However, this study shows that patients' motivation, capability, and social and environmental factors change over time. A systematic review of mHealth interventions for patients with depression supports the finding that patients' engagement with interventions changes over time [53], perhaps because their treatment needs and goals change over time. These changing needs of patients from their mHealth app interventions and their impact on their use behavior is further supported by another study conducted among patients with chronic illnesses [18]. Thus, we need to consider how our technologies can adapt to how users evolve over time.

In the current chronic care paradigm, the task of goal management is often left to the patients [3,54]. Our study findings highlight that those discussions regarding goal-oriented care are a one-time occurrence for study participants, which was facilitated by introducing the ePRO app. After setting goals with patients, providers often leave it up to patients to be responsible for their own goals. By contrast, patients do not bring up the topic of goals in their discussion as they perceive that their providers "are too busy" to attend to patients' goals and providers' time could be better spent on other condition-related concerns. This study finding reflects that there is a need for an ongoing conversation between the patient and provider about patient-centered goals to ensure that the goals and associated devices and tools are appropriate for the patient's needs and serve the purpose that the goal or device set out to do. Similarly, the interview data suggest that patients considered that their providers' enthusiasm for the ePRO intervention was important and influenced their interest in two ways: (1) monitoring of patient data by providers, which was considered important, and (2) providers' encouragement to keep using the ePRO app [55]. This finding highlights the need for further education and training tools for health care providers on how to effectively have a goal-oriented conversation with patients and within interprofessional teams [10,56].

Strengths

The descriptive qualitative approach of this research allowed us to identify multiple social-behavioral factors that influenced patients' enrollment in the study and subsequent discontinuation

or continuation. In addition, by using a restorying method, the findings were interpretive, allowing for the identification of nuanced patterns and interrelationships between identified themes. Furthermore, the longitudinal timeline of the study (15 months) allowed us to explore the factors that contribute to patients' use behavior in the long term, which is underexplored in the current literature [15]. Finally, as the SCT by Bandura [22] has been widely used to explore an individual's behavior and actions toward health-enhancing behavior, we were able to compare the findings of this study with previous literature [22,33,43,57]. For example, previous studies have identified that patients' self-efficacy, motivation, capacity, social and environmental influences, and perceived consequences affect their use behavior of an mHealth app.

Limitations

Owing to scheduling conflicts or loss to follow-up of participants, we were not able to interview all of them at either time point. As a result, a potential limitation of the study is that those who participated in the interviews may be unique as compared with those who chose not to. However, the sample size was too small to assess whether the difference between the 2 groups was significant. However, the interviews that were conducted were in-depth and provided rich information. Furthermore, the patient population represented in this study was recruited from only 3 of the 6 FHTs involved in this study. It is possible that some additional findings may have been obtained by looking across all 6 sites. However, the sample in this study represented 59% (22/37) of the total participants in the larger study. As is the case with case study research, it is also possible that findings may not be transferable to other models of primary care such as community health centers or solo practice environments. Furthermore, the participant demography suggests that the study patient population was less complex and well resourced, meaning that, on average, patients had a low number of chronic conditions and high income and educational attainment levels, which might not be representative of general complex patients. Therefore, the findings of this study may not be transferable to patients living in resource-poor communities or who have lower income or education levels. In addition, the underrepresentation of low-income individuals is a common occurrence across multiple research studies and requires attention in study design to facilitate this population's participation [58].

Conclusions

In many cases, mHealth or any health innovation will have expected impacts if people use it as intended. To better predict, explain, and increase the actual use of innovations, we need to understand why different target user groups continue or discontinue the use of an innovation. This study identifies that multilevel factors contribute to complex patients' decision to continue or discontinue using a goal-oriented app. In addition, our findings show that there is a need for ongoing, productive patient-provider interactions to set and modify patients' goals according to their changing health and social needs. Future research should consider patients' social and behavioral contexts when implementing mHealth apps and similar technological interventions for complex patients.

Acknowledgments

This study was funded by the Canadian Institute for Health Research eHealth Innovation Partnerships Program (CIHR–143559). The funders did not have any involvement in the study design and execution, analyses and interpretation of the data, and review or approval of the manuscript for publication.

Authors' Contributions

FT contributed to data collection, writing of the first draft, reviewing, and editing. TA contributed to data collection and analysis, writing of the first draft, reviewing, and editing. BM, ML, SWM, and RU contributed to writing (critical review) and editing. CSG conceived and designed the study and contributed to data collection, writing of the initial draft, reviewing and editing, and supervision.

Conflicts of Interest

BM holds a paid consultancy with the Scottish Government to provide advice on remote patient monitoring. However, BM has no ownership stake in the electronic Patient-Reported Outcome app. Therefore, we do not foresee any conflict of interest.

Multimedia Appendix 1
Additional information about data analysis.
[DOCX File , 32 KB - humanfactors_v9i4e37684_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3041 KB - humanfactors_v9i4e37684_app2.pdf]

References

1. Lewis J, Ray P, Liaw S. Recent worldwide developments in eHealth and mHealth to more effectively manage cancer and other chronic diseases - a systematic review. Yearb Med Inform 2016 Nov 10(1):93-108 [FREE Full text] [doi: 10.15265/IY-2016-020] [Medline: 27830236]
2. Putranto D, Rochmawati E. Mobile applications for managing symptoms of patients with cancer at home: a scoping review. Int J Nurs Pract 2020 Aug;26(4):e12842. [doi: 10.1111/ijn.12842] [Medline: 32347599]
3. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
4. Rosella LC, Fitzpatrick T, Wodchis WP, Calzavara A, Manson H, Goel V. High-cost health care users in Ontario, Canada: demographic, socio-economic, and health status characteristics. BMC Health Serv Res 2014 Oct 31;14:532 [FREE Full text] [doi: 10.1186/s12913-014-0532-2] [Medline: 25359294]
5. Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. Prev Med Rep 2018 Dec;12:284-293 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
6. Schaink AK, Kuluski K, Lyons RF, Fortin M, Jadad AR, Upshur R, et al. A scoping review and thematic classification of patient complexity: offering a unifying framework. J Comorb 2012;2:1-9 [FREE Full text] [doi: 10.15256/joc.2012.2.15] [Medline: 29090137]
7. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
8. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth Enhanced Chronic Care Model: a theory derivation approach. J Med Internet Res 2015 Apr 01;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
9. Steele Gray C, Barnsley J, Gagnon D, Belzile L, Kenealy T, Shaw J, et al. Using information communication technology in models of integrated community-based primary health care: learning from the iCOACH case studies. Implement Sci 2018 Jun 26;13(1):87 [FREE Full text] [doi: 10.1186/s13012-018-0780-3] [Medline: 29940992]
10. Gordon K, Steele Gray C, Dainty KN, DeLacy J, Ware P, Seto E. Exploring an innovative care model and telemonitoring for the management of patients with complex chronic needs: qualitative description study. JMIR Nurs 2020;3(1):e15691. [doi: 10.2196/15691] [Medline: 34345777]
11. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019 Oct 10;5:2055207619882181 [FREE Full text] [doi: 10.1177/2055207619882181] [Medline: 31656632]
12. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. BMC Med 2015 Apr 23;13(1):91. [doi: 10.1186/s12916-015-0279-6]
13. Gammon D, Johannessen LK, Sørensen T, Wynn R, Whitten P. An overview and analysis of theories employed in telemedicine studies. A field in search of an identity. Methods Inf Med 2008;47(3):260-269. [Medline: 18473093]
14. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
15. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
16. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013 Aug 30;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
17. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
18. Yin K, Jung J, Coiera E, Laranjo L, Blandford A, Khoja A, et al. Patient work and their contexts: scoping review. J Med Internet Res 2020 Jun 02;22(6):e16656 [FREE Full text] [doi: 10.2196/16656] [Medline: 32484449]
19. Tsai C. Integrating social capital theory, social cognitive theory, and the technology acceptance model to explore a behavioral model of telehealth systems. Int J Environ Res Public Health 2014 May 07;11(5):4905-4925 [FREE Full text] [doi: 10.3390/ijerph110504905] [Medline: 24810577]
20. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
21. Serlachius A, Schache K, Kieser A, Arroll B, Petrie K, Dalbeth N. Association between user engagement of a mobile health app for gout and improvements in self-care behaviors: randomized controlled trial. JMIR Mhealth Uhealth 2019 Aug 13;7(8):e15021 [FREE Full text] [doi: 10.2196/15021] [Medline: 31411147]
22. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]

23. Steele Gray C, Chau E, Tahsin F, Harvey S, Loganathan M, McKinstry B, et al. Assessing the implementation and effectiveness of the electronic patient-reported outcome tool for older adults with complex care needs: mixed methods study. J Med Internet Res 2021 Dec 02;23(12):e29071 [FREE Full text] [doi: 10.2196/29071] [Medline: 34860675]
24. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
25. Steele Gray C, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
26. Steele Gray C, Gravesande J, Hans PK, Nie JX, Sharpe S, Loganathan M, et al. Using exploratory trials to identify relevant contexts and mechanisms in complex electronic health interventions: evaluating the electronic patient-reported outcome tool. JMIR Form Res 2019 Feb 27;3(1):e11950 [FREE Full text] [doi: 10.2196/11950] [Medline: 30810532]
27. Tahsin F, Tracy S, Chau E, Harvey S, Loganathan M, McKinstry B, et al. Exploring the relationship between the usability of a goal-oriented mobile health application and non-usage attrition in patients with multimorbidity: a blended data analysis approach. Digit Health 2021 Oct 05;7:20552076211045579 [FREE Full text] [doi: 10.1177/20552076211045579] [Medline: 34868614]
28. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
29. Glazier RH, Redelmeier DA. Building the patient-centered medical home in Ontario. JAMA 2010 Jun 02;303(21):2186-2187. [doi: 10.1001/jama.2010.753] [Medline: 20516421]
30. Singh H, Tahsin F, Nie JX, McKinstry B, Thavorn K, Upshur R, et al. Exploring the perspectives of primary care providers on use of the electronic Patient Reported Outcomes tool to support goal-oriented care: a qualitative study. BMC Med Inform Decis Mak 2021 Dec 29;21(1):366 [FREE Full text] [doi: 10.1186/s12911-021-01734-0] [Medline: 34965860]
31. Focus on Geography Series, 2016 Census. Statistics Canada. 2017. URL: https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Index-eng.cfm [accessed 2022-06-14]
32. Middelweerd A, Te Velde SJ, Mollee JS, Klein MC, Brug J. App-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults (Active2Gether): descriptive study of the development and content. JMIR Res Protoc 2018 Dec 21;7(12):e185 [FREE Full text] [doi: 10.2196/resprot.7169] [Medline: 30578198]
33. Kim E, Han S. Determinants of continuance intention to use health apps among users over 60: a test of social cognitive model. Int J Environ Res Public Health 2021 Oct 01;18(19):10367 [FREE Full text] [doi: 10.3390/ijerph181910367] [Medline: 34639666]
34. Grace-Farfaglia P. Social cognitive theories and electronic health design: scoping review. JMIR Hum Factors 2019 Jul 19;6(3):e11544 [FREE Full text] [doi: 10.2196/11544] [Medline: 31325290]
35. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160 [FREE Full text] [doi: 10.1080/08870440512331317670]
36. Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theory Into Practice 2000 Aug;39(3):124-130. [doi: 10.1207/s15430421tip3903_2]
37. Creswell J, Poth CN. Qualitative Inquiry and Research Design Choosing Among Five Approaches. Thousand Oaks, California: SAGE Publications; 2016.
38. Ollerenshaw JA, Creswell JW. Narrative research: a comparison of two restorying data analysis approaches. Qual Inquiry 2016 Jun 29;8(3):329-347. [doi: 10.1177/10778004008003008]
39. Beauchamp M, Crawford K, Jackson B. Social cognitive theory and physical activity: mechanisms of behavior change, critique, and legacy. Psychol Sport Exercise 2019 May;42(5):110-117. [doi: 10.1016/j.psychsport.2018.11.009]
40. Lincoln Y, Guba E. Establishing trustworthiness. In: Naturalistic Inquiry. Thousand Oaks, California: SAGE Publications; 1985.
41. Bradway M, Pfuhl G, Joakimsen R, Ribu L, Grøttland A, Årsand E. Analysing mHealth usage logs in RCTs: explaining participants' interactions with type 2 diabetes self-management tools. PLoS One 2018;13(8):e0203202 [FREE Full text] [doi: 10.1371/journal.pone.0203202] [Medline: 30161248]
42. Vancouver JB, Thompson CM, Williams AA. The changing signs in the relationships among self-efficacy, personal goals, and performance. J Appl Psychol 2001 Aug;86(4):605-620. [doi: 10.1037/0021-9010.86.4.605] [Medline: 11519645]
43. Brown LA, Wiley JF, Wolitzky-Taylor K, Roy-Byrne P, Sherbourne C, Stein MB, et al. Changes in self-efficacy and outcome expectancy as predictors of anxiety outcomes from the CALM study. Depress Anxiety 2014 Aug;31(8):678-689 [FREE Full text] [doi: 10.1002/da.22256] [Medline: 24619599]

44. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
45. Gucciardi E, Demelo M, Offenheim A, Stewart DE. Factors contributing to attrition behavior in diabetes self-management programs: a mixed method approach. BMC Health Serv Res 2008 Feb 04;8:33 [FREE Full text] [doi: 10.1186/1472-6963-8-33] [Medline: 18248673]
46. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 06;22(1):e15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]
47. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
48. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
49. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Canadian Psychology / Psychologie canadienne 2008 Aug;49(3):182-185. [doi: 10.1037/a0012801]
50. Greenhalgh T, Wherton J, Sugarhood P, Hinder S, Procter R, Stones R. What matters to older people with assisted living needs? A phenomenological analysis of the use and non-use of telehealth and telecare. Soc Sci Med 2013 Sep;93:86-94 [FREE Full text] [doi: 10.1016/j.socscimed.2013.05.036] [Medline: 23906125]
51. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021 Dec 16;9(12):e29098 [FREE Full text] [doi: 10.2196/29098] [Medline: 34927597]
52. Choudhury A, Asan O, Choudhury MM. Mobile health technology to improve maternal health awareness in tribal populations: mobile for mothers. J Am Med Inform Assoc 2021 Oct 12;28(11):2467-2474 [FREE Full text] [doi: 10.1093/jamia/ocab172] [Medline: 34459478]
53. Molloy A, Anderson PL. Engagement with mobile health interventions for depression: a systematic review. Internet Interv 2021 Dec;26:100454 [FREE Full text] [doi: 10.1016/j.invent.2021.100454] [Medline: 34621626]
54. May C. The Burden of Treatment and the Burden of Symptoms: the problem of workload and institutional boundaries in chronic disease. Working Papers in Health Sciences. URL: https://www.southampton.ac.uk/assets/centresresearch/documents/wphs/CRM%20the%20Burden%20of%20Treatmen.pdf [accessed 2021-05-06]
55. Hanley J, Fairbrother P, McCloughan L, Pagliari C, Paterson M, Pinnock H, et al. Qualitative study of telemonitoring of blood glucose and blood pressure in type 2 diabetes. BMJ Open 2015 Dec 23;5(12):e008896 [FREE Full text] [doi: 10.1136/bmjopen-2015-008896] [Medline: 26700275]
56. Steele Gray C, Grudniewicz A, Armas A, Mold J, Im J, Boeckxstaens P. Goal-oriented care: a catalyst for person-centred system integration. Int J Integr Care 2020 Nov 04;20(4):8 [FREE Full text] [doi: 10.5334/ijic.5520] [Medline: 33199976]
57. Williams KE, Bond MJ. The roles of self-efficacy, outcome expectancies and social support in the self-care behaviours of diabetics. Psychol Health Med 2010 Aug 19;7(2):127-141. [doi: 10.1080/13548500120116076]
58. Joseph G, Kaplan CP, Pasick RJ. Recruiting low-income healthy women to research: an exploratory study. Ethn Health 2007 Nov;12(5):497-519 [FREE Full text] [doi: 10.1080/13557850701616961] [Medline: 17978946]

Abbreviations

ePRO: electronic Patient-Reported Outcome
FHT: Family Health Team
mHealth: mobile health
SCT: Social Cognitive Theory

Edited by A Kushniruk; submitted 02.03.22; peer-reviewed by J Lin, K Uludag, A Choudhury, M Witry, T Patel; comments to author 14.05.22; revised version received 13.09.22; accepted 06.10.22; published 30.11.22.

Please cite as:
Tahsin F, Austin T, McKinstry B, Mercer SW, Loganathan M, Thavorn K, Upshur R, Steele Gray C
Examining Use Behavior of a Goal-Supporting mHealth App in Primary Care Among Patients With Multiple Chronic Conditions: Qualitative Descriptive Study
JMIR Hum Factors 2022;9(4):e37684
URL: https://humanfactors.jmir.org/2022/4/e37684
doi:10.2196/37684
PMID:36449335

Original Paper

Low– and Medium–Socioeconomic-Status Group Members' Perceived Challenges and Solutions for Healthy Nutrition: Qualitative Focus Group Study

Sander Hermsen[1], PhD; Alex van Kraaij[1], MSc; Guido Camps[1,2], PhD, DVM

[1]OnePlanet Research Center, Wageningen, Netherlands

[2]Wageningen University and Research, Wageningen, Netherlands

Corresponding Author:
Sander Hermsen, PhD
OnePlanet Research Center
Bronland 10
Wageningen, 6708 WH
Netherlands
Phone: 31 317 791 009
Email: sander.hermsen@imec.nl

Abstract

Background: Although digital tools for healthy nutrition have shown great potential, their actual impact remains variable as digital solutions often do not fit users' needs and barriers. This is especially poignant for priority communities in society. Involving these groups in citizen science may have great benefits even beyond the increase in knowledge of the lives and experiences of these groups. However, this requires specialized skills. Participants from priority groups could benefit from an approach that offers sensitization and discussion to help them voice their needs regarding healthy nutrition and technology to support healthy eating.

Objective: This study aimed to gather insights into people's thoughts on everyday eating practices, self-regulation in healthy eating, and skill acquisition and on applying technological innovations to these domains.

Methods: Participants answered 3 daily questionnaires to garner their current practices regarding habits, self-regulation, skills, and technology use surrounding healthy eating and make it easier for them to collect their thoughts and experiences (sensitization). Within a week of filling out the 3 questionnaires, participants took part in a web-based focus group discussion session. All sessions were transcribed and analyzed using a thematic qualitative approach.

Results: A total of 42 people took part in 7 focus group interviews of 6 people each. The analysis showed that participants would like to receive support from technology for a broad range of aspects of nutrition, such as measuring the effect their personal nutrition has on their individual health, providing them with reliable product information, giving them practical guidance for healthy eating and snacking, and reducing the burden of registering food intake. Technology should be easy to use, reduce burdens, and be tailored to personal situations. Privacy and cost were major concerns for the participants.

Conclusions: This study shows that people from low– and medium–socioeconomic-status groups have a need for specific support in tailoring their knowledge of healthy nutrition to their own situation and see technology as a means to achieve this.

(JMIR Hum Factors 2022;9(4):e40123) doi:10.2196/40123

KEYWORDS

nutrition; citizen science; socioeconomic status; digital technology

Introduction

Background

Nutrition is an important element of health throughout our life, from the earliest stages of infancy [1] to old age [2]. Malnutrition [3], overconsumption [4], and unbalanced diet composition [5] are strongly associated with a broad range of debilitating health conditions. The proliferation of digital technology among the population in the last decade has enabled the development and use of a broad range of health-oriented software, mobile apps, and wearable products capable of supporting us in healthy behavior change [6,7]. In digital technology for health, nutrition is one of the most frequently

targeted behavioral domains [7,8]. Although digital tools for healthy nutrition have shown great potential to improve health outcomes, enhance patient experience in health care, and control or reduce costs, their actual impact remains variable. One of the main reasons for this is the fact that digital solutions often offer an approach that does not fit users' needs and barriers, putting too much effort into knowledge attainment and too little into dealing with everyday practices and habits and social and cultural aspects of nutrition [9].

The reduced efficacy of digital health innovations is especially visible in and poignant for priority populations and communities in society [10]—people with lower income, less education, and less high-status employment than more advantageous groups in society and often referred to as vulnerable or low socioeconomic status (SES) in scientific research [11-13]. These groups require priority in digital health research and practice as their quality-adjusted life expectancy is, on average, almost a decade of healthy years lower than that of high-SES groups. This prioritization is hindered by the fact that, in scientific literature, little is known about the barriers, needs, and desires surrounding nutrition for low-SES groups in daily life [14-16].

Citizen Science Approaches

Involving priority groups in citizen science—a research approach in which citizens themselves produce reliable scientific knowledge [17]—has great benefits beyond the increase in knowledge of the lives and experiences of these groups. The underrepresentation of individuals from priority groups [18] has known consequences on the scientific and innovative outcomes of research projects, such as interventions that work well for high-SES groups but less so for priority groups [19-21], with the current level of evidence insufficient to inform policy and practice [21]. Involving priority groups improves the chances of developing products and interventions that match the needs and possibilities of these groups. Furthermore, taking part in citizen science approaches can empower low-SES group members by increasing their skills and knowledge and providing them with a platform to share their experiences [22,23].

However, making the most of citizen science takes knowledge and awareness of those phenomena that the research looks into as well as the opportunity and skills to develop research questions and designs, make observations, apply scientific reasoning, and make sense of discoveries [24]. Unfortunately, this proves difficult for most people and even more so for people in priority groups. They could benefit from an approach that empowers them to take part in citizen science projects. Such an approach should sensitize people to the subject matter to make it easier for them to collect their thoughts and experiences. Furthermore, the approach should foster discussion to help people formulate and bring across their ideas, needs, and barriers and share and compare them with others.

Such a process, in which people from priority groups take part in discussions to help them voice their needs and desires regarding digital technology for healthy nutrition, has as yet not taken place. Research into needs and barriers regarding technology for healthy living has until now only looked at specific target groups within the population, such as older adults [25] or parents with young children [26]. These studies mainly looked at prerequisites of using technology within a given use case, such as weight management; no studies asked the target groups which use cases would interest them in the first place. However, the available literature [27-36] can give us a hint of what themes we might expect to emerge when we ask people about what they would find important in technological innovations aimed at a healthy lifestyle, such as dealing with everyday eating practices—dietary tracking, weight management, mindful eating, balanced nutrition, and eating habits; self-regulation processes (eg, in snacking behavior); and support in cooking and shopping skills.

Objectives

Therefore, this paper had the aim of gathering insights into barriers, needs, desires, and use cases that people, especially those from priority backgrounds, experience regarding technology for healthy nutrition. To do so, the study focused on people's thoughts on applying technological innovations to everyday eating practices, self-regulation (eg, snacking), and skill acquisition. The approach presented in this study aimed to empower priority group members to contribute to the development of research agendas in this area.

Methods

Overview

To gather insights into the barriers, needs, and desires surrounding healthy nutrition and technology for healthy eating, we performed a 2-phase qualitative study. In the first phase, participants answered 3 daily questionnaires to garner their current practices regarding habits, self-regulation, skills, and technology use surrounding healthy eating. Furthermore, the questionnaires served as a "sensitizer" to support participants in thinking about their healthy eating desires and needs. In the second phase, participants took part in focus group discussion sessions in which they discussed their challenges, experiences, and perceived solutions regarding healthy eating. During the focus group sessions, a short presentation on technology for health at the host institute served as further sensitization to foster further discussion on the potential of technology for supporting healthy nutrition. Participants filled out the questionnaires in the week starting on Monday, November 9, 2020; all focus group sessions were held between November 12, 2020, and November 17, 2020.

Participants

The study aimed to include people living in the Dutch province of Gelderland and able to speak and understand Dutch. Furthermore, we aimed to include as many people from low-SES groups as possible. Indicators of low SES [37] were education (general secondary education or secondary vocational level as the highest attainment), income (<€18,390 [US $18,160.30] per annum), and employment (unemployed or otherwise). This proportion of low-SES participants should be at least as high as or higher than the low-SES prevalence in the population (ie, 20%-35% of participants). A priori sample size calculations in qualitative research are subject to conceptual debate and practical uncertainty. Saturation (ie, the moment when adding more data does not lead to new insights) is often seen as a

criterion for the inclusion of more participants once the analysis has started. As a rule of thumb, 20 to 40 participants are usually considered sufficient to achieve saturation [38]. Therefore, to deal with dropout, only to be expected when conducting research under pandemic restrictions [39], we aimed to include 40 participants.

Participants were recruited from a large panel of potential participants provided by a field research agency in the Netherlands. The research team invited all members of this panel who met the eligibility criteria: living in Gelderland and a maximum educational level of secondary or vocational education. Those panel members who opted to participate were then contacted by phone to explain the study procedures. Participants received an incentive of €50 (US $49.38) to spend at a Dutch web store for taking part. All participants signed an informed consent form digitally before taking part in the study.

Data Collection

Questionnaires

Participants were asked to fill out a daily questionnaire for 3 consecutive days. The goal of the questionnaires was 2-fold. First, they served as a means of gathering data on participant attitudes and behaviors surrounding healthy nutrition. Second, and more importantly, they served as a "sensitizer" to trigger participants to think about their healthy eating desires and needs. The concept of "sensitizing" is derived from participatory design practices and aims to help people think about their habits and needs in preparation for creative sessions [40]. Each questionnaire had its own theme. The first questionnaire featured questions on participant demographics, regular meals, and eating habits. The second questionnaire contained questions on self-regulatory aspects of eating behavior—drinking and snacking behavior. The third and final questionnaire contained questions on skills relevant to healthy nutrition: purchasing (healthy) food, preparing food, dealing with waste and leftovers, and using technology for healthy nutrition. The 3 questionnaires are included in Multimedia Appendix 1.

The questionnaires were sent by email at 7 PM; reminder emails were sent at 9 PM in case the questionnaire had not been filled out yet. In case participants still had not filled out the questionnaire by the next morning, they received a phone call from the supporting research company to remind them. Each questionnaire took approximately 15 to 30 minutes to fill out. Questions could be answered by selecting the relevant value from a Likert scale or list of options. Each questionnaire ended with an open question in which participants could freely describe their needs regarding the theme of the day.

For reasons of concision and because the main aim of the questionnaires was to trigger thinking about nutrition and technology in the participants rather than gather insights on nutrition habits and behaviors of the participants, the results of the questionnaires about eating habits, snacking, and drinking are not reported in the main body of this paper. An overview of these results is available in Multimedia Appendix 2. Only the results on technology use for healthy nutrition are reported in the *Results* section of this paper.

Focus Group Sessions

Within a week of filling out the 3 questionnaires, participants took part in one of 7 focus group discussion sessions. Owing to the restrictions because of the COVID-19 pandemic, the discussion sessions took place on the web using Microsoft Teams. Before the session, participants received the invitation for the discussion session and technical instructions for joining by email. Participants provided their permission for video recording of the sessions for analysis purposes.

Each of the 7 discussion sessions had room for 5 to 7 participants and was scheduled to last 60 minutes. A researcher served as session host; furthermore, one other researcher and an assistant took part. The session host started the session with a brief general introduction and a video and sound check for every participant. Participants were free to use the "raise hand" button to ask questions or comment at any time. The session host guided the conversation by assigning speaking turns to each participant.

The interviews were semistructured; each participant answered 5 predefined questions. In between answers, researchers and other participants could ask additional questions or comment on the provided answer. The first 3 questions were as follows: "What would you still like to know about nutrition, considering the different aspects of nutrition covered in the questionnaires?" "What would you like to change when it comes to nutrition, considering the different aspects of nutrition covered in the questionnaires?" and "How could technology support you in your challenges related to nutrition, considering the different aspects of nutrition covered in the questionnaires?" After this segment, the second researcher presented 3 examples of technological solutions to nutritional challenges: a "smart" toilet seat, a "smart" wristband, and an ingestible measurement device. Participants were then given the opportunity to ask questions or provide comments afterward. Questions 4—"What new thoughts, ideas or directions for technological solutions for nutritional challenges come to mind after hearing these examples?"—and 5—"What other aspects in daily life would you like to see more technological support, besides nutrition?"—were then discussed, and the session was concluded. The session leaders gave no further prompts or hints regarding potential themes between the main questions.

Data Processing and Analysis

Questionnaires

All demographic data from the questionnaires (eg, age, gender, and education of the participants) were (pseudo-)anonymized (eg, age was divided into age groups of 18-39 years, 40-54 years, 55-64 years, and ≥65 years) and read in using R (R Foundation for Statistical Computing) [41]. For each question with Likert scales, averages and ranges were calculated. Answers to open questions were coded using the method and coding scheme derived from the discussion session analysis (see the following section).

Focus Group Sessions

The research team manually transcribed the recordings of the sessions. They anonymized the transcript by removing personal

information. All transcripts were then read into qualitative analysis software [42] and analyzed using a method based on thematic analysis with both deductive and inductive components [43] such that insights from theory and evidence guided the analysis but, at the same time, we were open to novel themes that emerged during the analysis. Following this approach [44,45], 2 researchers (AvK and SH) first performed a primary analysis using an initial coding scheme based on expected themes derived from theory and previous literature on 29% (2/7) of the session transcripts individually and then compared their codings to ascertain similar interpretations. They then applied inductive coding to identify themes and patterns in the data that were not yet covered by the coding scheme. A further iteration of the analysis then took place to ascertain the confidence in the deductive codings. The coding scheme was then modified to better reflect emergent themes, and all relevant text segments were coded again. This step was repeated until no more issues arose.

The initial coding scheme consisted of 3 main themes, all derived from literature on determinants of healthy eating. The first theme revolved around everyday practices and habitual patterns in nutrition: meal contents, meal settings, and other regular eating-related behaviors. The second theme revolved around controlling impulses such as snacking, and the third theme revolved around skills needed for purchasing, preparing, storing, and postprocessing food. For each theme, the initial coding scheme consisted of the subthemes knowledge, desires, nontechnological solutions, and technological solutions for healthy nutrition. The illustrative quotes presented in the *Results* section were abbreviated for length and clarity.

How Might We Statements

To translate the results of the focus group discussions into insights suitable to inform a research agenda, we made use of *How Might We* statements [46]. This method, also known as "How-Tos," consists of rephrasing statements about challenges or desires for healthy nutrition in such a way that they support idea generation.

Ethics Approval

This study was exempt from approval from ethical committees under Dutch and European regulations. Under Dutch regulations, as there is no burden on the participants in this type of study (solely interview-based), it requires no ethics approval by the medical ethical committee [47]. Before enrollment, all participants provided written informed consent for the collection and use of data.

Results

Participants

A total of 42 people took part in 7 focus group interviews of 6 people each. In total, 38% (16/42) of the participants described their gender as man, and 62% (26/42) described it as woman. Of the 42 participants, 7 (17%) were in the 18 to 39 age group, 14 (33%) were in the 40 to 54 age group, and the remaining 21 (50%) were in the 55 to 64 age group. A total of 2% (1/42) of the participants reported their BMI as underweight (<18 kg/m^2), 43% (18/42) were at a healthy weight (<25 kg/m^2), 40% (17/42) were overweight (25-30 kg/m^2), and 14% (6/42) were severely overweight (>30 kg/m^2). Almost half (20/42, 48%) of the participants met at least one criterion of low SES, 5% (2/42) met all 3 criteria (education, profession, and income), 17% (7/42) met 2 criteria (education and profession, education and income, or profession and income), and 26% (11/42) met 1 criterion (education, profession, or income). In total, 52% (22/42) of the participants met none of the criteria; see Table 1 for an overview of low SES indicators in the participants. Of the 42 participants, 11 (26%) reported living alone, 2 (5%) lived with their partner, 15 (36%) lived with their children but without a partner, and 14 (33%) lived with their partner and children. A total of 48% (20/42) of the participants indicated some sort of health issue, with stomach and gut complaints (5/42, 12%) and type 2 diabetes (2/42, 5%) being the most frequent. All participants came from the Netherlands and were Dutch.

Table 1. Overview of low socioeconomic status (SES) indicators (N=42).[a]

Characteristic	Participants, n (%)
Household gross annual income	
<€14,100[b] (US $13,923.90)	*2 (5)*
€14,100 to €29,500 (US $13,923.90-$29,131.60)	*7 (17)*
€29,500 to €36,000 (US $29,131.60-$35,550.40)	8 (19)
€36,000 to €43,500 (US $35,550.40-$42,956.80)	7 (17)
€43,500 to €73,000 (US $42,956.80-$72,088.40)	8 (19)
€73,000 to €87,100 (US $72,088.40-$86,012.30)	5 (12)
>€87,100 (US $86,012.30)	3 (7)
No answer	2 (5)
Education	
None or primary only or language courses only	*1 (2)*
Lower and secondary vocational education	*5 (12)*
Higher levels of secondary education	7 (17)
Old-style vocational education (<1989)	29 (69)
Profession	
Employed in government	2 (5)
Employed in governmental institutes	5 (12)
Employed in a company	16 (38)
Self-employed	3 (7)
Housewife or househusband	*1 (2)*
Incapacitated	*10 (24)*
Unemployed, job seeking, or social assistance	*4 (10)*
Other	1 (2)

[a]Of the 42 participants, 2 (5%) met all 3 criteria (education, profession, and income), 7 (17%) met 2 criteria (education and profession, education and income, or profession and income), and 11 (26%) met 1 criterion (education, profession, or income). In total, 52% (22/42) of the participants met none of the criteria and, therefore, were categorized as medium-SES.

[b]Categories in italics are indicators of low SES.

Questionnaires

A large majority (38/42, 90%) of the participants indicated that they knew examples of technology to help them with healthy nutrition, be they apps, websites, or wearable devices. People who knew none were evenly distributed between the low- and medium-SES participant groups (2/42, 5% each). A total of 60% (12/20) of the participants from the low-SES group indicated using at least one technology-based solution to help with aspects of healthy nutrition, as did 64% (14/22) of the participants from the medium-SES group. Participants' familiarity with and use of different categories of technological solutions are listed in Table 2.

Table 2. Knowledge and use of technology for healthy nutrition in low– and medium–socioeconomic status (SES) participants. Participants could select more than 1 category; "none of the above" excluded all other categories (N=42).

	I know of, n (%)		I already use, n (%)	
	Low SES	Medium SES	Low SES	Medium SES
Apps, websites, or smart devices to measure intake	12 (29)	8 (19)	2 (5)	1 (2)
Apps, websites, or smart devices to help prevent waste	5 (12)	8 (19)	0 (0)	0 (0)
Apps, websites, or smart devices to measure calorie need and use	11 (26)	12 (29)	4 (10)	7 (17)
Apps, websites, or smart devices to help prepare meals	9 (21)	9 (21)	10 (24)	14 (33)
Apps, websites, or smart devices to help with shopping	6 (14)	3 (7)	6 (14)	4 (10)
None of the above	2 (5)	2 (5)	8 (19)	8 (19)
Low-technology solutions (eg, pen and paper to register food)	N/A[a]	N/A	5 (12)	5 (12)

[a]N/A: not applicable.

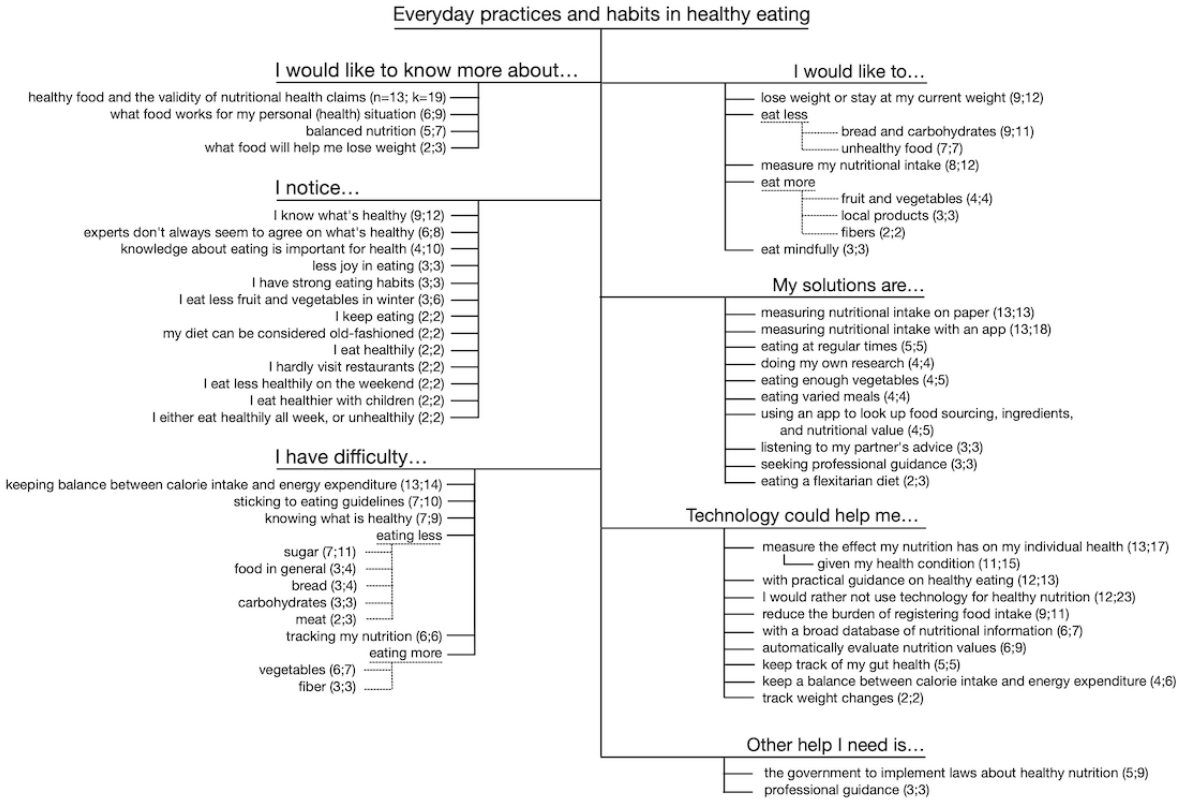
Focus Group Sessions

Overview

The original coding scheme consisted of the following main themes: everyday practices and habits in healthy eating; self-regulation of snacking and drinking; and skills in purchasing, cooking, and storing food. These main themes also emerged from the qualitative analysis. In total, 3 new main themes also emerged: user experience, privacy, and cost of technological solutions for healthy nutrition. For each of the original main themes, the original subthemes—knowledge, desires, nontechnological solutions, and technological solutions—were insufficient to reveal the structure in the retrieved codes. A new categorization of subthemes emerged: *I would like to know...*, *I notice...*, *I have difficulty with...*, *I wish...*, *My solutions are...*, *Technology could help me with...*, and *Other help I could use is...*

Thematic saturation [48] occurred during the analysis, achieving both code saturation, with the code group structure already present after coding 29% (2/7) of the focus group interview transcripts, and meaning saturation, with all important themes developed after 71% (5/7) of the transcripts (Figure 1). Stratification of results between participants from the low- and medium-SES groups revealed that statements for every main theme occurred more or less equally between both groups, and no differences occurred.

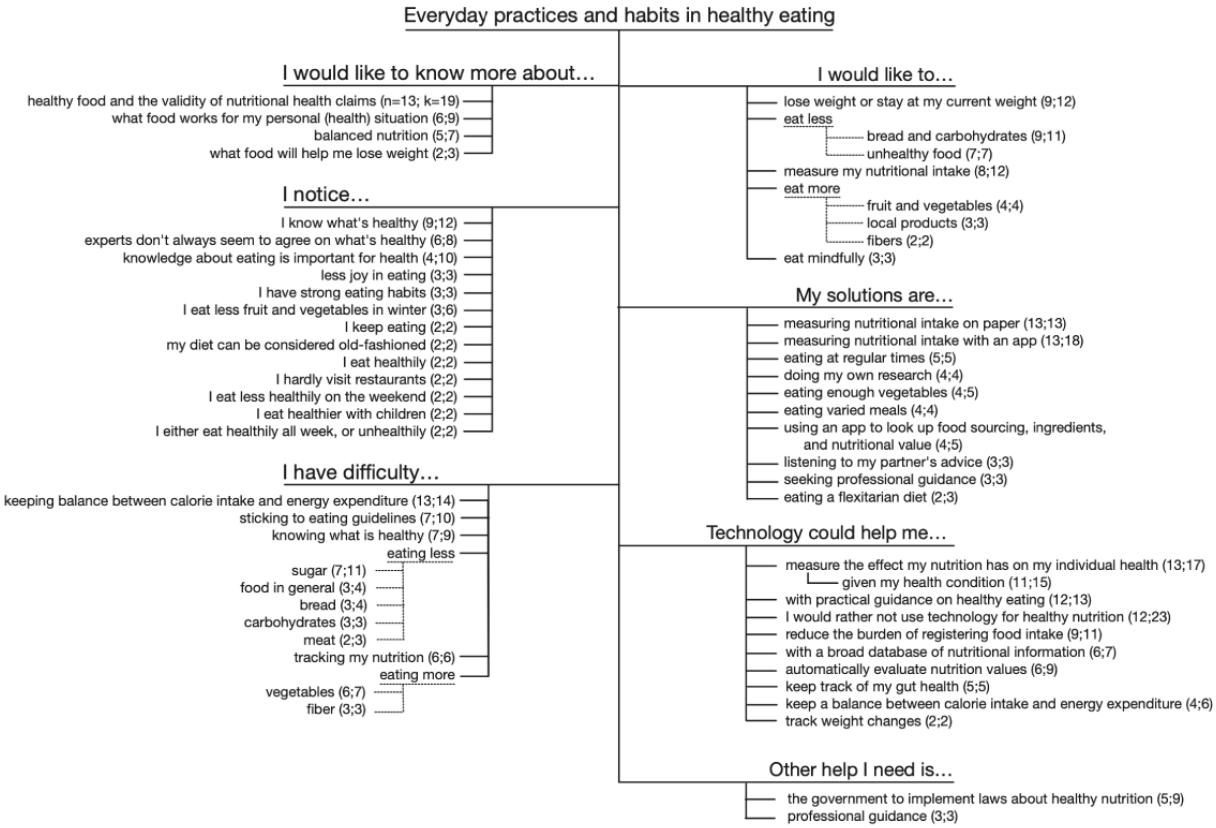
Figure 1. Code saturation. For each main theme and session, the occurrence of statements with codes within that theme is displayed. Largest squares: >15 mentions; medium-sized squares: 6 to 15 mentions; smallest squares: 1 to 5 mentions. k: number of unique segments about this sub-theme; n: number of unique participants mentioning this sub-theme.



Everyday Practices and Habits in Healthy Eating

The first major theme emerging from the data concerned nutritional habits, recurring behaviors, and everyday practices in healthy eating. This theme received the most attention in the discussion sessions, with more than one-third (633/1797, 35.23%) of all the unique coded segments being within this theme. Figure 2 shows its code structure.

Figure 2. Code structure for the theme Everyday practices and habits in healthy eating. k: number of unique segments about this sub-theme; n: number of unique participants mentioning this sub-theme.



The analysis revealed a tension between a widespread opinion in the participants that they were well aware of what constitutes healthy nutrition on the one hand and (often at the same time) difficulties in knowing what is healthy on the other. Participant 1, for instance, stated both that they "Actually [...] feel [they] don't really need to know much more about healthy food" and that they "could do with more information and tips and tricks on healthy eating." Other subthemes within this theme appear to be related to this discrepancy—people would like to know more about the validity of health claims and notice that experts do not always seem to agree:

> *I have trouble getting the right information, one person says butter is good for you, the other says it's not.* [Participant 40]

> *In science, there often seem to be competing findings. One study says that elderly people need less protein, another says you need to make sure you get enough protein. That makes it hard to do the right thing in my situation.* [Participant 17]

A total of 14% (6/42) of the participants expressed a wish for technology to help them determine nutritional truth by providing a broad database of nutritional information. In total, 29% (12/42) of the participants would like technology that helps them with practical guidance on healthy eating:

> *[a solution for] at home, when you take a piece of gingerbread, you can scan it, and the app tells you that you haven't had your daily allowance and tells you to "go ahead, take the gingerbread, with a cup of tea."* [Participant 28]

> *For instance, an app that I can tell "I've already eaten [this food] today and I'm still peckish," and the app tells me to "just eat this, then you'll be fine."* [Participant 10]

A second major subtheme concerned personalized nutrition and knowing what effect a personal diet has on their individual health. In total, 31% (13/42) of the participants expressed a general desire for technology to evaluate the impact of nutrition on their individual health. Moreover, 26% (11/42) of the participants mentioned this in the context of their health condition. This makes technological innovations to foster personalized nutrition a desire of more than half of the participants (24/42, 57%):

> *I put a lot of energy into healthy nutrition, but I have reached my limits in what I can do myself, also with regard to my diabetes. I would like to know more about that.* [Participant 48]

> *I sometimes have heartburn for days at a time, and I ask myself what I shouldn't have eaten...Too many*

> *spices? I would really like to find out why this is.* [Participant 41]
>
> *I have two health issues that contradict each other when it comes to nutrition. I do need help for that.* [Participant 31]

A related subtheme was keeping track of the balance between calorie intake and energy expenditure. A total of 31% (13/42) of the participants mentioned this as something that they struggled with, and 10% (4/42) expressed a wish to address this balance using technology:

> *I would like to combine the personal information of a smartwatch with food intake, that would be great. It does need to be accessible though, so the older generation [refers to self] can still deal with it.* [Participant 49]

To make sense of the way nutrition affects their personal situation, more than half of the participants (26/42, 62%) already tracked their food intake in one way or another. In total, 31% (13/42) of the participants used pen-and-paper solutions to do so, and another 31% (13/42) used an app. Many participants (11/42, 26%) expressed difficulties they encountered with food registration methods; when tracker apps or other pen-and-paper methods were mentioned, these mentions were often combined with the burden these solutions posed on the user (which is also a recurrent theme in the *User Experience* section):

> *I have tried to track [my food intake] with my activity tracker app, but that took me all day, at least it felt like that. You need to enter everything precisely and that's just too much for me.* [Participant 1]

A total of 21% (9/42) of the participants mentioned a wish for technology that takes away the burden of food tracking, and another 14% (6/42) of the participants mentioned a wish for technology that automatically assesses nutritional values. In total, 29% (12/42) of the participants mentioned that they would rather not use technology at all when it comes to healthy eating but, of these 12 participants, 6 (50%) still expressed wishes for technological solutions to measure the effect of nutrition on their personal situation:

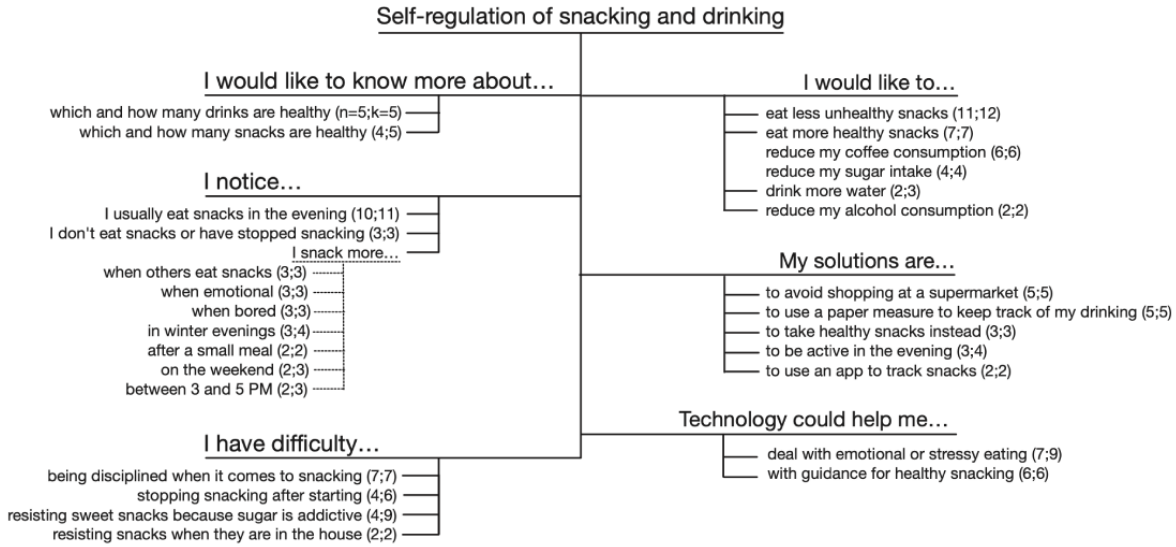
> *I am not an app-using person, I don't really look up things and I don't really care about it. [...] But I would like a watch or something that just tells me I'm getting too heavy or have been eating the wrong things.* [Participant 41]

Other subthemes that emerged were weight loss and maintaining a healthy weight, avoiding undesirable ingredients and foods, eating healthier food in general, and keeping track of gut health; participants would like to see technological solutions that offer practical guidance to do so. Beyond technology, 12% (5/42) of the participants would like to see extended governance in healthy eating, and 7% (3/42) would like to receive professional guidance.

Self-regulation

The second main theme concerned self-regulation in snacking and drinking behavior, the latter mostly concerning alcohol and sweetened beverages. With 197 unique segments, this theme was less pronounced than the previous one, but the analysis revealed that this is still an issue that many people struggle with. The code structure for this theme is shown in Figure 3.

Figure 3. Code structure for the theme Self-regulation of snacking and drinking. k: number of unique segments about this sub-theme; n: number of unique participants mentioning this sub-theme.



Dealing with snacking was a recurrent issue in this theme. Participants would like to know which and how many snacks could be considered healthy; noticed that there were certain times and conditions in which they snacked more; had difficulty with snacking discipline; and would like to snack less and, if they did, more healthily:

> *I would like to snack less when I'm bored, especially in the evenings.* [Participant 30]
>
> *I don't really snack in between meals, unless I'm having a binge when I'm not feeling good about myself.* [Participant 41]

> *I consider myself an emotional eater, for instance when I am sewing, I always need something nice to take the tension away. It would help me to get notified when my stress level is getting too high.* [Participant 26]

Regarding technology, they would mostly like interventions to help them deal with the effect of stress and emotions on snacking (7/42, 17% of the participants) and guidance on healthy snacking (6/42, 14% of the participants):

> *Maybe you could develop an app that makes you happy with healthy food, without needing sugars. Because when you're down, you tend to take chocolate and stuff, and then you notice you get tired because of the sugar. Maybe you can make some kind of mind app?* [Participant 45]

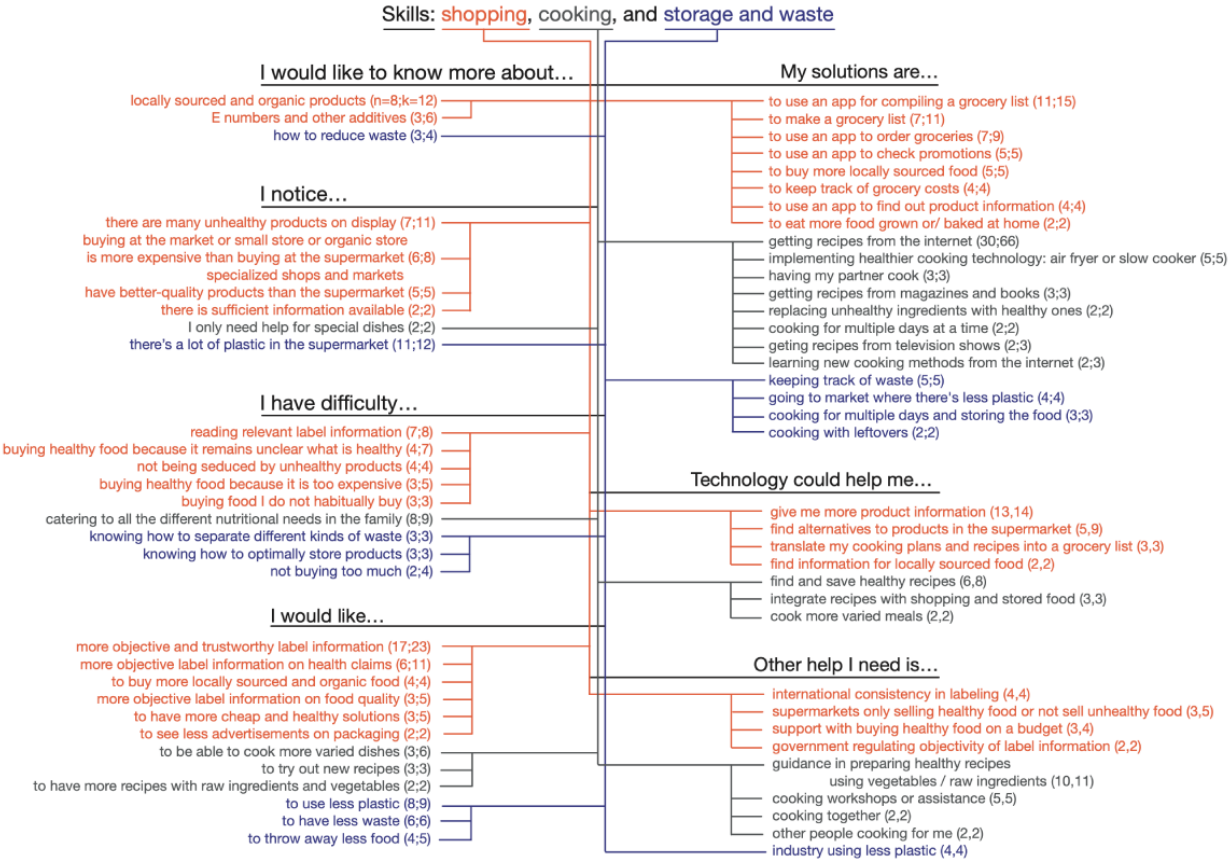
> *I know what stress does to your gut health, so I could definitely see myself using [technology] there.* [Participant 49]

Drinking turned out to be less relevant to the participants; they would like to know which and how many drinks could be considered healthy, drink more water, and reduce alcohol and coffee consumption. None of the participants expressed a wish to use technology to track or alter drinking behavior.

Skills: Shopping, Cooking, and Waste and Storage

The third main theme concerned the skills and information a person needs for healthy nutrition: acquiring healthy food, cooking it in a palatable manner, storing the food before preparation, and dealing with leftovers and waste. A total of 663 unique segments were included in this theme, with shopping (n=345, 52%) making up the largest part and cooking (n=201, 30.3%) and waste and storage (n=117, 17.6%) less so. The code structure for this theme is shown in Figure 4.

Figure 4. Code structure for the theme Skills: shopping, cooking, and storage and waste. k: number of unique segments about this sub-theme; n: number of unique participants mentioning this sub-theme.



Regarding shopping for groceries, the largest subtheme concerned labeling information. A total of 17% (7/42) of the participants found it difficult to interpret the relevant information on the labels, and 10% (4/42) thought that what was healthy remained unclear. In total, 40% (17/42) of the participants would like to see more objective and trustworthy label information, and another 14% (6/42) would like to see more objective information on health claims:

> *There are so many products in stores, and the ingredients are listed in such small print, you just cannot read what is in there. [...] And there are so many different names for [for example] sugar. It's like they are fooling you.* [Participant 7]

> *Sometimes you buy something and it says on the packaging that it's healthy, with some sort of label too, and then you take a closer look and it turns out the food has many, many calories, and sugars, or aspartame which I'm not a fan of at all.* [Participant 27]

In total, 7% (3/42) of the participants particularly would like to know more about health claims regarding additives. Technology, participants thought, could help them with providing neutral product information (13/42, 31% of the participants) and find alternatives to products in the supermarket (5/42, 12% of the participants):

> *I always take the same, for example, potatoes, and maybe other kinds are simply better, but I don't know that. And an app could send you a notification that I could replace the product I just took with something healthier.* [Participant 48]

> *I would like an app that can do a quick scan of the products in the store, and can give me advice on what to buy and what not, for instance because of the sugar content.* [Participant 21]

A second recurrent subtheme was a wish for more knowledge of (8/42, 19% of the participants) and more opportunities to buy (4/42, 10% of the participants) locally sourced and organic food. Unfortunately, even though these products are often seen as being of better quality, they are seen as more expensive. Regarding technology, 5% (2/42) of the participants would like to know more about where to find locally sourced and organic food:

> *I would like an app that tells me what locally sourced produce is available at this moment. So, if there's a farmer harvesting tomatoes this week, it tells me this is a regionally sourced alternative that's available.* [Participant 42]

Regarding cooking, 2 main subthemes emerged. First, many participants (8/42, 19%) expressed difficulty with catering to the different nutritional needs of family members:

> *I cook for different people, me and my husband, my children, and an elderly person. So, I always need to pay attention to what I cook, so everybody gets the nutrition that fits their life phase. That combination is really difficult.* [Participant 28]

> *My children do a lot of sports, and I have a sedentary profession and a tendency to snack. [...] You just can't give everybody the same food.* [Participant 45]

A second subtheme was preparing varied and healthier dishes:

> *I mainly cook pasta now. [...] It's just too seductive because it's quick. I'm by myself and I don't like cooking so I want something quick.* [Participant 41]

With regard to technology, most participants (30/42, 71%) used the internet as a source of recipes. Participants wanted technology to enable them to find and save healthy recipes and integrate those recipes with shopping and stored food, especially for more varied meals. This also connected to similar wishes regarding technology for shopping:

> *I would like an app that combines recipes with supermarket shopping lists, so you don't have to enter everything twice.* [Participant 47]

> *[when you find a recipe] there's an entire shopping list that you need to store somewhere and take out in the supermarket. It would be so much easier if you just can do it in one go: a cooking plan combined with a shopping list.* [Participant 39]

Regarding dealing with waste and storage, reducing waste, especially plastics, was the issue participants were mostly interested in (8/42, 19%). Only 2% (1/42) of the participants had a use for technology in this theme—an app that reminded them of leftovers.

Criteria for Use: User Experience, Privacy, and Cost

The analysis revealed 3 themes related to criteria for use: user experience, privacy, and cost. In total, this theme consisted of 198 unique segments, with 98 (49.5%) related to user experience, 46 (23.2%) concerning privacy, and 24 (12.1%) concerning financial aspects of technology for healthy nutrition.

Concerning user experience, 33% (14/42) of the participants mentioned the high burden of use that current technological solutions for healthy eating place on the user. They would welcome innovations that reduce this burden:

> *All those apps are way too complicated. You need to keep track of so many things for just a small result. That's just not feasible.* [Participant 28]

> *I would look at certain apps, but it's so much trouble to keep it up and do all the data entry that I just leave it be. I just cannot persevere.* [Participant 16]

> *I used one of those apps once, and I needed to fill in all kinds of data per day. But then I couldn't find one [foodstuff] and then I thought, now it's no longer accurate for me, so forget this. I never returned to the app.* [Participant 30]

Therefore, ease of use was mentioned by 36% (15/42) of the participants as imperative for adoption:

> *I want it to be without the hassle of needing to set things up; I should just have to push a button and it would tell me what I have eaten this morning and what I still need.* [Participant 5]

> *It needs to be easy, that's right. I won't walk away from technology, but it needs to serve me, not require me to pay constant attention and add data. It needs to support me, not the other way around.* [Participant 22]

In total, 14% (6/42) of the participants also wanted to ascertain that the advice they received from technology was tailored to their personal situation:

> *As long as I don't get my husband's or my son's advice, because they are thin as planks, they can eat anything and not gain a gram. For me it's the opposite. What if I get his advice, like, eat another eight sandwiches, that's not going to work, is it?* [Participant 49]

> *Often [the advice] is aimed at people who do a lot of sport, not people like me who just want to move a bit more.* [Participant 3]

Other concerns were efficacy (4/42, 10% of the participants), reliability and validity (2/42, 5% of the participants), safety

(3/42, 7% of the participants), and connectivity with other technological solutions (2/42, 5% of the participants).

In total, 55% (23/42) of the participants talked about privacy concerns with technology for healthy nutrition. Participants wanted to know who they were sharing their data with:

> *I don't mind sharing things, for instance with you [researchers]. At least then the data will be beneficial for everybody. But I do want to know what is shared.* [Participant 1]

> *Even when they say they won't use the data, I am still scared they will.* [Participant 19]

They wanted to have control over who obtains the data:

> *I want it to be designed in a way that it's protected. That I can share the data with someone else because I allow it, for instance with your general practitioner if you have health concerns. And it needs to be safe, because in this day and age, with all the hackers and information leaks and data that gets sold on, nobody needs yet another device that will bring your entire life out in the open.* [Participant 31]

Participants were also well aware of the risks and negative consequences of data sharing:

> *What if an insurance company rejects me based on this data?* [Participant 5]

> *...you hear a lot about, for instance, a smart fridge, that others can hack into that and use it in some way or other.* [Participant 19]

A final recurring concern was cost. A total of 26% (11/42) of the participants mentioned high cost as a barrier to the adoption of technology for healthy nutrition:

> *To get comprehensive information, you often need a paid version, and that stops me.* [Participant 21]

> *I would be interested if it were covered by my health insurance. But that's a problem in itself, because when everybody wants this technology, the insurance fees need to go up and people with limited budgets can't afford it anymore. One way or another, it needs to be paid for.* [Participant 25]

> *I don't have any apps because I don't dare touch them, for fear of them costing me money.* [Participant 7]

> *Smartwatches, fitbits, yes I've seen them, but I just cannot afford them.* [Participant 9]

Implications for Research Agenda: How Might We Statements

To facilitate the translation of the insights from the focus group discussions into statements that could serve as starting points for future research, we reformulated the wishes participants expressed for technological support in healthy nutrition as *How Might We* statements. The full list of statements is presented in Table 3.

Table 3. Overview of the How Might We statements derived from participants' expressed needs for technology support for healthy nutrition (N=42).

	Participants, n (%)
How might we...	
...measure the effect individual nutrition has on a person's unique health situation?	24 (57)
...provide people with detailed product information when shopping?	13 (31)
...give practical guidance for healthy and varied eating?	12 (29)
...reduce the burden of registering food intake?	9 (21)
...help people deal with emotional or stressy eating?	7 (17)
...provide a broad database of nutritional information?	6 (14)
...automatically evaluate nutrition values?	6 (14)
...give practical guidance for healthy snacking?	6 (14)
...help people find and save healthy recipes?	6 (14)
...help people keep track of their gut health?	5 (12)
...help people find products and alternatives for products in the supermarket?	5 (12)
...help people keep a balance between calorie intake and energy expenditure?	4 (10)
...help people translate cooking plans and recipes into grocery lists?	3 (7)
...help people integrate recipes with shopping and food stored at home?	3 (7)
...help people track weight changes?	2 (5)
...provide information on locally sourced food?	2 (5)
...help people cook more varied meals?	2 (5)
In such a way that...	
...it reduces the burden of keeping track of nutrition and reminding oneself of goals?	23 (55)
...the innovation is easy to use?	17 (40)
...it is tailored to people's personal situation?	6 (14)
...is effective?	4 (10)
...is safe?	3 (7)
...is reliable and valid?	2 (5)
...is connected to other solutions?	2 (5)

Discussion

Principal Findings

This study aimed to gather insights into use cases, barriers, and needs for technology in people from priority backgrounds to support them in healthy nutrition. To do so, participants from lower- and medium-SES groups filled out questionnaires and took part in a focus group discussion meeting. The study showed that participants would like to receive support from technology to measure the effect their personal nutrition has on their individual health; provide them with reliable product information; give them practical guidance for healthy eating and snacking; reduce the burden of registering food intake; help them deal with emotional or stress-based snacking; automatically evaluate nutritional values of their food intake; help them find and save healthy, varied recipes and translate these recipes, combined with stored food at home, into shopping lists; help them keep track of gut health; help them find alternatives to unhealthy products; help them keep a balance between calorie intake and energy expenditure; help them track weight changes; and provide information on locally sourced food. Technology should be easy to use; reduce the burden of keeping track of nutrition and intervening in healthy eating; and be tailored to their personal situation, effective, safe, reliable, valid, and connected to other "smart" solutions. Privacy and cost were major concerns for the participants.

The results show that tensions exist in the way people perceive and act upon healthy eating. On the one hand, participants indicated that they had all the knowledge they needed for healthy nutrition but, by contrast, (the same) participants told us that they were often confused by contradictory health claims and had difficulty putting their (often abstract) knowledge into everyday practice. A similar tension can be found in technology use. More than one-quarter of the participants (12/42, 29%) told us that they would rather refrain from using technology for healthy eating but, at the same time, had clear opinions on how technology could still help them. This tension could very well be driven by the experience participants have with current solutions, which are often burdensome and insufficiently tailored to their personal needs and situations. Personalized feedback

and guidance was the number-one use case described by the participants, especially regarding dealing with often complex health issues and different needs in the family. However, not every barrier, concern, or need described by the participants was accompanied by a wish for a technological solution. Abundance of plastic packaging was mentioned by many participants (11/42, 26%) as a concern, but none of them felt the need for digital health innovations to deal with this issue.

The aim of this study was to reach participants from lower- and medium-SES groups. The results showed that 45% (19/42) of the participants met at least one marker of membership of these groups. None of the remaining participants were from high-status groups. This makes this study successful, but caution regarding the generalizability of the results remains necessary. First, the markers describing low SES are vague and apply to a wide range of profiles, from people without formal education but with successful careers to people who have received higher education but are currently unemployed. This vagueness in definitions is not unique to this research but applies to all studies dealing with priority populations [37]. Second, selection bias is likely to have occurred in the recruitment of the participants. To take part in the study, participants had to be members of a specific research panel and express their interest in the study. This means that our results represent the opinions and interests of a convenience sample, and people without an interest in technology are not likely to have applied for the study. However, as these people are also most likely to be early adopters of technology to support them in healthy eating, the results of this study are still relevant to inform further research and innovation even if some selection bias occurred. Third, researcher bias, in which the interests of the researchers affect question order, question content, and interview procedures, can have occurred. All authors are professionally involved in research into and the development of technological innovations, which may have influenced the research. However, as 80% of innovations fail within 2 years of launch because of limited connection to everyday practices and barriers of users [49], this involvement also entails a sincere interest in what people do not find interesting or engaging in technology or when technology fails in supporting people in changing their behavior. Overall, the study managed to engage different voices than is usually the case, with some caveats given the potential for bias. Further research can support or refute the conclusions drawn in this paper.

The questionnaire results showed that the participants in this study can be seen as typically Dutch eaters, with sandwiches for breakfast and lunch and a hot meal for dinner often consisting of potatoes, vegetables, and meat [50]. This shows that we managed to engage the "common person" in this research but, of course, it limits the extent to which these results can be generalized to other diets and cultures. In addition to nutritional content, Dutch eating culture sees nutrition as an individual matter or, at the most, a matter of the nuclear family [51]. Many (if not most) other cultures lay more emphasis on social practices of eating [51]. It is to be expected, but open to future research, that people from these cultures would place more value on solutions that cater to different needs within the (extended) family or enhance social or festive aspects of healthy eating [52-55].

The desire for personalization of healthy eating corresponds with recent evidence (see the study by König et al [28] for a review); technological interventions for healthy nutrition need to be customizable and tailored to individual needs. Usability issues [56-60], privacy concerns [58,61], and cost concerns [57,62] are also well known as barriers to adoption and sustained engagement with technological innovations.

This study used a novel method derived from participatory design. Therefore, the study can serve as an example of a case study of citizen science approaches that go beyond simply having citizens gather data for preset research questions. The approach described in this paper is a first step toward the development of a well-defined method for the first stages of participatory research [63], in which citizens have the opportunity to voice their thoughts, concerns, and desires about what is researched. In the follow-up stages, the research questions generated by the approach described in this paper can then be explored by the citizens involved in the discussion and others. This study shows the promise of this approach as it produced a broad range of research questions that could be further explored. The sensitizer phase, using questionnaires to help shape people's thoughts about subjects such as habitual eating, self-regulation, and skills, did its work in that participants seemed well prepared to take part in the discussion. Sensitizing people to subjects could potentially carry with it a risk of social desirability bias in questionnaire and interview responses. However, this bias is mitigated by a better preparedness in the participant—they have already had the opportunity to form their own opinion on the subject and are less likely to be subject to answering biases during the session [64].

The sensitizing exercise performed during the discussion session, in which we showed the participants examples of innovations, was much less successful. After the presentation, hardly any new themes emerged. The exercise did not seem to help participants think about technology beyond what they already knew and used, as evidenced by the fact that the ideas for solutions that the participants mentioned were mostly similar to the apps and websites that they already used. A more immersive approach could be more fruitful in this case, such as "provotyping" [65,66]. In this approach, a combination of "provocation" and "prototyping," participants receive gentle and safe "provocations" through experiencing (potential) innovations that could help them deal with the subject matter. Often, such "provotypes" lead to discussing latent cultural norms and taboos, automatic behavioral patterns such as deeply entrenched habits, and other processes and attitudes that normally escape conscious scrutiny. Ideally, people should experience the prototype for an extended period, not just as part of a discussion session.

The research delivered a range of research questions (reflected in the list of *How Might We* statements) on the barriers, needs, desires, and use cases people from priority backgrounds experience regarding technology for healthy nutrition. These statements have already proved useful to help inform the discussion on the research agenda on technology for healthy

nutrition at the host institution. They helped shape conversations with potential partners from academia and industry, and this approach is now used in other scientific programs within the organization. However, the establishment of a research agenda is often complex and must align the needs and demands of practice (industry and health care), science, and funding agencies. Furthermore, a technology at an advanced readiness level must be available for further development. All this entails that, in practice, the impact of citizens' input on research agendas will remain smaller than desired. Nonetheless, this approach can make sure that the voice of the citizen is considered, which in itself is already a great step forward.

Conclusions

This study provided an overview of challenges, needs, and barriers that people from low- and medium-SES groups see when it comes to healthy nutrition. The results show that these people, even though they think of themselves as having knowledge of what constitutes healthy eating, are in need of specific support when it comes to knowing what is healthy for their specific situation and specific support for changing everyday eating practices and habits and obtaining skills needed for healthy eating. The study also showed how technology can play a role in supporting these people and that usability, privacy, and cost need to be considered. Finally, the study provided an approach to help people from priority groups voice their needs and concerns and can serve as a blueprint to use input from these groups to inform research and development agendas.

Acknowledgments

The authors would like to thank Joris van Nieuwstadt and Gert Jan van Amerongen for their contribution to the provotyping exercise in the discussion sessions and I&O Research for providing access to their participant panel. OnePlanet Research Center is supported by funding from the province of Gelderland.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

The study was conceived and designed by SH, AvK, and GC. AvK led the discussion groups with support from OnePlanet colleagues. All analyses were conceived by all authors and performed by AvK and SH. The first draft of the paper was written by SH. All authors reviewed the paper, made key intellectual contributions to the content and reporting, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensitizing questionnaires.
[PDF File (Adobe PDF File), 116 KB - humanfactors_v9i4e40123_app1.pdf]

Multimedia Appendix 2
Eating behavior questionnaire overview.
[PDF File (Adobe PDF File), 57 KB - humanfactors_v9i4e40123_app2.pdf]

References

1. Stephenson J, Heslehurst N, Hall J, Schoenaker DA, Hutchinson J, Cade JE, et al. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. Lancet 2018 May;391(10132):1830-1841. [doi: 10.1016/s0140-6736(18)30311-8]
2. Govindaraju T, Sahle B, McCaffrey T, McNeil J, Owen A. Dietary patterns and quality of life in older adults: a systematic review. Nutrients 2018 Jul 26;10(8):971 [FREE Full text] [doi: 10.3390/nu10080971] [Medline: 30050006]
3. Saunders J, Smith T. Malnutrition: causes and consequences. Clin Med (Lond) 2010 Dec 01;10(6):624-627 [FREE Full text] [doi: 10.7861/clinmedicine.10-6-624] [Medline: 21413492]
4. Samartı´n S, Chandra RK. Obesity, overnutrition and the immune system. Nutrition Res 2001 Jan;21(1-2):243-262. [doi: 10.1016/s0271-5317(00)00255-4]
5. Parillo M, Riccardi G. Diet composition and the risk of type 2 diabetes: epidemiological and clinical evidence. Br J Nutr 2007 Mar 09;92(1):7-19. [doi: 10.1079/bjn20041117]
6. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Human Behav 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]

7. Šmahel D, Macháčková H, Šmahelová M, Čevelíček M, Almenara C, Holubčíková J. Using mobile technology in eating behaviors. In: Digital Technology, Eating Behaviors, and Eating Disorders. Cham: Springer; 2018.
8. Taj F, Klein MC, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
9. Ludden G, Hermsen S. Healthy eating and behaviour change. In: Design for Wellbeing. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2019.
10. Sinha C, Schryer-Roy AM. Digital health, gender and health equity: invisible imperatives. J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii1-ii5 [FREE Full text] [doi: 10.1093/pubmed/fdy171] [Medline: 30329082]
11. Munari SC, Wilson AN, Blow NJ, Homer CS, Ward JE. Rethinking the use of 'vulnerable'. Aust N Z J Public Health 2021 Jun 05;45(3):197-199. [doi: 10.1111/1753-6405.13098] [Medline: 33818873]
12. Flaskerud JH, Winslow BJ. Conceptualizing vulnerable populations health-related research. Nurs Res 1998;47(2):69-78. [doi: 10.1097/00006199-199803000-00005] [Medline: 9536190]
13. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, et al. Socioeconomic status in health research: one size does not fit all. JAMA 2005 Dec 14;294(22):2879-2888. [doi: 10.1001/jama.294.22.2879] [Medline: 16352796]
14. Ballering C, Schreurs H, Renders C, Kooiker S, van Ameijden E. Een inkijk in verhalen achter leefstijlgewoontes. Tijds gezondheids wetenschappen 2013 Jun 29;91(5):263-269. [doi: 10.1007/s12508-013-0094-4]
15. Inglis V, Ball K, Crawford D. Why do women of low socioeconomic status have poorer dietary behaviours than women of higher socioeconomic status? A qualitative exploration. Appetite 2005 Dec;45(3):334-343. [doi: 10.1016/j.appet.2005.05.003] [Medline: 16171900]
16. Wardle J, Steptoe A. Socioeconomic differences in attitudes and beliefs about healthy lifestyles. J Epidemiol Community Health 2003 Jun;57(6):440-443 [FREE Full text] [doi: 10.1136/jech.57.6.440] [Medline: 12775791]
17. Irwin A. Citizen Science A Study of People, Expertise and Sustainable Development. London: Routledge; 1995.
18. Sorensen A, Jordan R, LaDeau S, Biehler D, Wilson S, Pitas JH, et al. Reflecting on efforts to design an inclusive citizen science project in West Baltimore. Citizen Sci Theor Pract 2019;4(1):13. [doi: 10.5334/cstp.170]
19. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci 2010 Jun;33(2-3):61-83; discussion 83. [doi: 10.1017/S0140525X0999152X] [Medline: 20550733]
20. Sarink D, Peeters A, Freak-Poli R, Beauchamp A, Woods J, Ball K, et al. The impact of menu energy labelling across socioeconomic groups: a systematic review. Appetite 2016 Apr 01;99:59-75. [doi: 10.1016/j.appet.2015.12.022] [Medline: 26723238]
21. South E, Rodgers M, Wright K, Whitehead M, Sowden A. Reducing lifestyle risk behaviours in disadvantaged groups in high-income countries: a scoping review of systematic reviews. Prev Med 2022 Jan;154:106916 [FREE Full text] [doi: 10.1016/j.ypmed.2021.106916] [Medline: 34922995]
22. Den Broeder L, Lemmens L, Uysal S, Kauw K, Weekenborg J, Schönenberger M, et al. Public health citizen science; perceived impacts on citizen scientists: a case study in a low-income neighbourhood in the Netherlands. Citizen Sci Theor Pract 2017 Nov 09;2(1):7. [doi: 10.5334/cstp.89]
23. Den Broeder L, Devilee J, Van Oers H, Schuit AJ, Wagemakers A. Citizen Science for public health. Health Promot Int 2018 Jun 01;33(3):505-514 [FREE Full text] [doi: 10.1093/heapro/daw086] [Medline: 28011657]
24. Wolf GI, De Groot M. A conceptual framework for personal science. Front Comput Sci 2020 Jun 30;2:21. [doi: 10.3389/fcomp.2020.00021]
25. Cabrita M, Op den Akker H, Tabak M, Hermens HJ, Vollenbroek-Hutten MM. Persuasive technology to support active and healthy ageing: an exploration of past, present, and future. J Biomed Inform 2018 Aug;84:17-30 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.010] [Medline: 29935348]
26. Burrows T, Hutchesson M, Chai LK, Rollo M, Skinner G, Collins C. Nutrition interventions for prevention and management of childhood obesity: what do parents want from an eHealth program? Nutrients 2015 Dec;7(12):10469-10479 [FREE Full text] [doi: 10.3390/nu7125546] [Medline: 26694456]
27. Stok FM, Hoffmann S, Volkert D, Boeing H, Ensenauer R, Stelmach-Mardas M, et al. The DONE framework: creation, evaluation, and updating of an interdisciplinary, dynamic framework 2.0 of determinants of nutrition and eating. PLoS One 2017 Feb 2;12(2):e0171077 [FREE Full text] [doi: 10.1371/journal.pone.0171077] [Medline: 28152005]
28. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
29. Choe E, Lee N, Lee B, Pratt W, Kientz J. Understanding quantified-selfers' practices in collecting and exploring personal data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; Apr 26 - May 1, 2014; Toronto Ontario Canada. [doi: 10.1145/2556288.2557372]
30. Sarcona A, Kovacs L, Wright J, Williams C. Differences in eating behavior, physical activity, and health-related lifestyle choices between users and nonusers of mobile health apps. Am J Health Educ 2017 Jul 11;48(5):298-305. [doi: 10.1080/19325037.2017.1335630]

31. Lieffers JR, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb;118(2):229-39.e3. [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
32. Almoraie NM, Saqaan R, Alharthi R, Alamoudi A, Badh L, Shatwan IM. Snacking patterns throughout the life span: potential implications on health. Nutr Res 2021 Jul;91:81-94 [FREE Full text] [doi: 10.1016/j.nutres.2021.05.001] [Medline: 34144310]
33. Marangoni F, Martini D, Scaglioni S, Sculati M, Donini LM, Leonardi F, et al. Snacking in nutrition and health. Int J Food Sci Nutr 2019 Dec 10;70(8):909-923. [doi: 10.1080/09637486.2019.1595543] [Medline: 30969153]
34. Hess J, Jonnalagadda S, Slavin J. What is a snack, why do we snack, and how can we choose better snacks? A review of the definitions of snacking, motivations to snack, contributions to dietary intake, and recommendations for improvement. Adv Nutr 2016 May;7(3):466-475 [FREE Full text] [doi: 10.3945/an.115.009571] [Medline: 27184274]
35. Engler-Stringer R. Food, cooking skills, and health: a literature review. Can J Diet Pract Res 2010 Sep;71(3):141-145. [doi: 10.3148/71.3.2010.141] [Medline: 20825697]
36. Caraher M, Lang T. Can't cook, won't cook: a review of cooking skills and their relevance to health promotion. Int J Health Promotion Educ 1999 Jan;37(3):89-100. [doi: 10.1080/14635240.1999.10806104]
37. Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health 2003 Apr;57(4):254-258 [FREE Full text] [doi: 10.1136/jech.57.4.254] [Medline: 12646539]
38. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol 2018 Nov 21;18(1):148 [FREE Full text] [doi: 10.1186/s12874-018-0594-7] [Medline: 30463515]
39. Suslow A, Giehl C, Hergesell J, Vollmar H, Otte I. Conducting qualitative research under pandemic restrictions – considerations, challenges, and benefits. Research Square. 2020. URL: https://www.researchsquare.com/article/rs-753390/v1 [accessed 2022-05-18]
40. Sanders E, Stappers PJ. Convivial Toolbox Generative Research for the Front End of Design. London, United Kingdom: Laurence King Publishing; 2012.
41. R: a language and environment for statistical computing. R Core Team. 2022. URL: https://www.R-project.org/ [accessed 2022-04-05]
42. MAXQDA 2020 – launching a new decade of research innovation. MAXQDA. URL: https://www.maxqda.com/new-maxqda-2020 [accessed 2022-04-05]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exercise Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
45. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014 Oct 16;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
46. Tassoul M. Creative Facilitation: A Delft Approach. Delft, NL: VSSD; 2005.
47. Wet medisch-wetenschappelijk onderzoek met mensen. Overheid nl. 2022 Nov 21. URL: https://wetten.overheid.nl/BWBR0009408/2022-07-01 [accessed 2022-11-21]
48. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar 26;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
49. Trott P. The role of market research in the development of discontinuous new products. Eur J Innov Manag 2001;4(3):117-126. [doi: 10.1108/14601060110390585]
50. van Rossum CT, Fransen HP, Verkaik-Kloosterman J, Buurma-Rethans EJ, Ocke MC. Dutch National Food Consumption Survey 2007-2010: diet of children and adults aged 7 to 69 years. Rijksinstituut voor Volksgezondheid en Milieu RIVM. 2011. URL: https://rivm.openrepository.com/handle/10029/261553 [accessed 2022-05-18]
51. Bukman AJ, Ronteltap A, Lebrun M. Interpersonal determinants of eating behaviours in Dutch older adults living independently: a qualitative study. BMC Nutr 2020 Nov 11;6(1):55 [FREE Full text] [doi: 10.1186/s40795-020-00383-2] [Medline: 33292680]
52. Danesi G. Pleasures and stress of eating alone and eating together among French and German young adults. Menu J Eating Hospitality Res 2012;1:77-91. [doi: 10.26048/5289-3f91]
53. Healy AE. Eating and ageing: a comparison over time of Italy, Ireland, the United Kingdom and France. Int J Comparative Sociol 2014 Dec 08;55(5):379-403. [doi: 10.1177/0020715214561132]
54. Rozin P. The meaning of food in our lives: a cross-cultural perspective on eating and well-being. J Nutrition Educ Behav 2005 Nov;37:S107-S112. [doi: 10.1016/s1499-4046(06)60209-1]
55. Danesi G. A cross-cultural approach to eating together: practices of commensality among French, German and Spanish young adults. Social Sci Infor 2017 Dec 11;57(1):99-120. [doi: 10.1177/0539018417744680]
56. Chan A, Kow R, Cheng JK. Adolescents' perceptions on smartphone applications (Apps) for health management. J Mob Technol Med 2017 Aug;6(2):47-55. [doi: 10.7309/jmtm.6.2.6]

57. Aljuraiban GS. Use of weight-management mobile phone apps in Saudi Arabia: a web-based survey. JMIR Mhealth Uhealth 2019 Feb 22;7(2):e12692 [FREE Full text] [doi: 10.2196/12692] [Medline: 30794205]
58. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
59. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
60. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
61. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
62. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
63. Haklay M. Citizen science and volunteered geographic information: overview and typology of participation. In: Crowdsourcing Geographic Knowledge. Dordrecht Netherlands: Springer; 2013.
64. Alvarado O, Storms E, Geerts D, Verbert K. Foregrounding algorithms: preparing users for co-design with sensitizing activities. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. 2020 Presented at: NordiCHI '20: Shaping Experiences, Shaping Society; Oct 25 - 29, 2020; Tallinn Estonia. [doi: 10.1145/3419249.3421237]
65. Boer L, Donovan J. Provotypes for participatory innovation. In: Proceedings of the Designing Interactive Systems Conference. 2012 Presented at: DIS '12: Designing Interactive Systems Conference 2012; Jun 11 - 15, 2012; Newcastle Upon Tyne United Kingdom. [doi: 10.1145/2317956.2318014]
66. Boer L, Donovan J, Buur J. Challenging industry conceptions with provotypes. CoDesign 2013 Jun;9(2):73-89. [doi: 10.1080/15710882.2013.788193]

Abbreviations

SES: socioeconomic status

Edited by A Kushniruk; submitted 07.06.22; peer-reviewed by C Mueller, K DiFilippo; comments to author 29.08.22; revised version received 15.09.22; accepted 06.10.22; published 02.12.22.

Please cite as:
Hermsen S, van Kraaij A, Camps G
Low– and Medium–Socioeconomic-Status Group Members' Perceived Challenges and Solutions for Healthy Nutrition: Qualitative Focus Group Study
JMIR Hum Factors 2022;9(4):e40123
URL: https://humanfactors.jmir.org/2022/4/e40123
doi:10.2196/40123
PMID:36459403

Original Paper

Preferences for a Mental Health Support Technology Among Chinese Employees: Mixed Methods Approach

Sijin Sun[1], MEng; Zheyuan Zhang[1], MA, MSc; Mu Tian[2], MSc; Celine Mougenot[1], PhD; Nick Glozier[3], PhD; Rafael A Calvo[1], PhD

[1]Dyson School of Design Engineering, Imperial College London, London, United Kingdom

[2]Luye Medical Group, Shanghai, China

[3]Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Corresponding Author:
Sijin Sun, MEng
Dyson School of Design Engineering
Imperial College London
Imperial College Rd
South Kensington
London, SW7 9EG
United Kingdom
Phone: 44 779967397
Email: s.sun20@imperial.ac.uk

Abstract

Background: Workplace mental health is under-studied in China, making it difficult to design effective interventions. To encourage the engagement with interventions, it is crucial to understand employees' motivation toward seeking help through technologies.

Objective: This study aimed to understanding how Chinese employees view digital mental health support technology and how mental health support technology could be designed to boost the motivation of Chinese employees to use it.

Methods: A mixed methods approach was used. In total, 458 Chinese employees (248/458, 54% female) in 5 industries (manufacturing, software, medical, government, and education) responded to a survey, and 14 employees and 5 managers were interviewed.

Results: Government data and employee responses showed that mental health support in China is limited. In the workplace, Chinese employees experience a lower sense of autonomy satisfaction compared with competence and relatedness. Although managers and employees try to empathize with those who have mental health issues, discrimination and the stigma of mental illness are rife in Chinese workplaces. Digital technologies are perceived as a potential medium for mental health interventions; however, privacy is a major concern.

Conclusions: The results of this study demonstrated the potential of self-help digital mental health support for Chinese employees. Interdisciplinary cooperation between design engineers and mental health researchers can contribute toward understanding the issues that engage or disengage users with digital mental health interventions.

(JMIR Hum Factors 2022;9(4):e40933) doi:10.2196/40933

KEYWORDS

mental health; digital health; workplace; China

Introduction

Over the last 2 decades, the link between workplaces and employees' mental health has become increasingly important to managers, employees, and their families worldwide [1]. In the United Kingdom, 15% of employees experience symptoms of existing mental health conditions [2]. In China, 35% of employees experience difficulties in stress management and 21% to 29% of the working population believe that their mental ill-being has a significant negative impact on their social life and working performance [3].

Despite these dire statistics, few people with mental disorders in China actively seek clinical support, with the rates being 17% in rural areas [4] and 25% in urban areas [5]. This has been

attributed to the stigma surrounding mental health, availability of adequate support, and cost of professional interventions. According to Zhang et al [5], 93.6% of the patients and caregivers in China misunderstand mental illnesses, preventing them from receiving appropriate help. There are only 0.036 clinical psychologists per 10,000 Chinese citizens, whereas the figure is 6 in the United States and 1.8 in the United Kingdom [6,7]. As for financial resources, the Chinese government has increased mental health expenditure per capita to US $0.13 by 2018, which is in contrast to the average of US $21.7 in Europe [8]. With the rapid increase in the direct medical costs of mental health, families in China face considerable financial burden, as out-of-pocket payments are the most common payment method for most mental health care expenditures [9].

According to the World Health Organization [10], mental health is defined as "a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community." As a country with rapid transition of economic and social system, the Chinese government is turning its attention to mental health–related issues and their negative impact on productivity [11].

Although programs and research on workplace mental health have grown in the West, occupational mental health has not been widely recognized as a significant research area in China. This has led to a lack of understanding of workplace mental illness, its sources, and the effectiveness of interventions. The understanding of these issues in Western workplaces may not generalize well to Chinese ones, given that the social perceptions of mental illness might be different; for example, patients with mental illness are traditionally considered a possible source of social unrest because of their potentially anomalous behavior [12]. The expectations about the relationship between employees and others from their companies would also differ, as in Chinese society, it is important to maintain *guanxi* (an underlying harmonious social relationship) with each other [13]. To design culturally acceptable digital interventions, a better understanding of the sociotechnical context is essential to meet the preferences and fulfill the needs of the target users. This includes understanding the "local definitions of personhood and the good life" and how they are situated [14].

In the West, digital well-being interventions have shown the potential to reduce the risks of mental illness in the workplace and support those who are already ill with positive acceptance [15]. Meta-analyses from 2017-2021 [16-18] showed that eHealth interventions can reduce the level of mental illness in employees and act as a treatment by reducing illness symptoms in those who are already ill. However, in China, although there are more than 986 million smartphone users [19] as of 2021, there are few studies on workplace digital mental health interventions. Meta-analyses in 2017 and 2019 [16,17] only included 1 study in China—a web-based mental health training program for university staff and students in Hong Kong by Mak et al [20].

In a 2019 study by Peking University Health Science Center [21], researchers conducted a qualitative assessment of all the existing smartphone-based mental health support apps from the iOS market and 3 major Android markets in China. In total, 63 unique apps were identified in the market following the criteria of (1) accessibility, (2) targeting the public (not designed for mental health experts), (3) technical robustness, and (4) evidence-based content. The apps were built on the basis of 2 primary features: 67% of the apps provide mental health education either by informative articles or by relevant courses (14 with free content and 28 with paid content); and 65% of the apps were built to provide videoconference counseling services. Only 7 of the 63 apps contained courses designed on the basis of verified psychological therapy or training. In addition, each app was measured using a Mobile App Rating Scale (MARS) [22] in 4 dimensions: engagement, function, esthetics, and information. Each MARS item uses a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent); the maximum scores for the Chinese apps were 4.0, 4.3, 4.0, and 4.0, respectively, much lower than the results of 4.6, 4.75, 4.83, and 4.6 using the same scale on mental well-being apps in the English language [21,22]. This suggests room for improvement in many aspects of the app design.

For the design of effective and acceptable digital mental health support technologies, it is crucial for design engineers to understand "How do Chinese employees view digital mental health support technology, and how could a mental health support technology be designed to boost the motivation of Chinese employees from using it?"

Methods

Study Structure Overview

A mixed method user research was conducted to provide insights into Chinese employees' understanding of mental health and their preferences for interventions, privacy, terms used, platform, and style. This paper discusses 2 studies: a quantitative survey-based study and a qualitative analysis of interviews with employees and managers.

Study 1: Survey

Sample and Recruitment

Employees were invited to the study via internal messaging tools in 5 Chinese local companies and organizations collaborating in the project: a manufacturing company (approximately 500 employees), software company (200 employees), state administration office (1400 officers), medical company (400 employees), and university department (50 students and staff). The factory of the manufacturing company was in a tier-3 city, where mental health resources are more limited [11] than the other 4 companies and organizations in tier-1 provincial capitals.

The anonymous invitation link was published in the companies' and organizations' group chat rather than directly sent to individuals to ensure that participation in this web-based survey was voluntary. Note that bias exists, as participants who are familiar with mental health and mental illness were more likely to voluntarily complete the survey; however, in the recruitment messages, participants were encouraged to complete the survey

even if they did not have previous experiences with mental well-being support. The link to the survey was open for 1 month.

Instead of emails, messaging tools such as "WeChat" and "DingTalk" are more frequently used for the communication between colleagues in China. To disseminate the survey, we chose to send messages in the company's group chat. The exact number of employees who read the messages in the group chat was unknown.

To be eligible, participants had to be Chinese residents, older than 18 years, and working in one of the 5 companies or organizations. Participants were provided with an information sheet describing the study and requested to sign a consent form before the start of the survey.

From the recruitment of participants to the end of the questionnaire collection, none of the cities, where the companies and organizations were based, were in COVID lockdown.

Survey Structure

The first study was conducted in the form of a web-based survey in simplified Chinese with WJX ("问卷星," a Chinese web-based survey platform). The survey took 10-15 minutes to complete, and its questions were organized into the following sections using standardized questionnaires when available:

1. Demographic information (gender, age, and role in the company) was measured using generic options provided by WJX.

2. History of mental health issue ("Are you currently or previously undergoing any form of psychological treatment?")

3. Smartphone use: internet use, current use, and satisfaction with apps for stress relief

4. Work motivation: psychological needs satisfaction (Basic Psychological Needs at Work Scale [BPNWS]) [23]

5. Mental well-being supportive preference: experience of and hindrance to receiving mental health support, experience of using and willingness to use different types of digital mental health support, preferred features for digital mental health support (based on the list of features recorded in a previous study for the Headgear app) [24]

6. Further participation (if the participant wanted to be contacted for the interview, they could leave their contact details).

Data Analysis

Statistical results collected from the backend database of WJX were first analyzed by industry and then aggregated to make an overall sample of the working population in China.

According to the self-determination theory [25], the satisfaction of the 3 basic psychological needs of autonomy, competence, and relatedness is positively related to psychological well-being and workplace engagement and negatively related to ill-being. In this part of the survey, the participants were invited to complete the BPNWS [23], which is a 12-item instrument on a 6-point scale. In a study by Brien et al [23], the French and English versions of the BPNWS were validated. The scale was translated into Chinese by 2 bilingual researchers. The Cronbach α coefficient was used to validate the translated version of the scale. One-way ANOVA was used to investigate the differences in basic psychological needs among the 5 industries.

Study 2: Interview

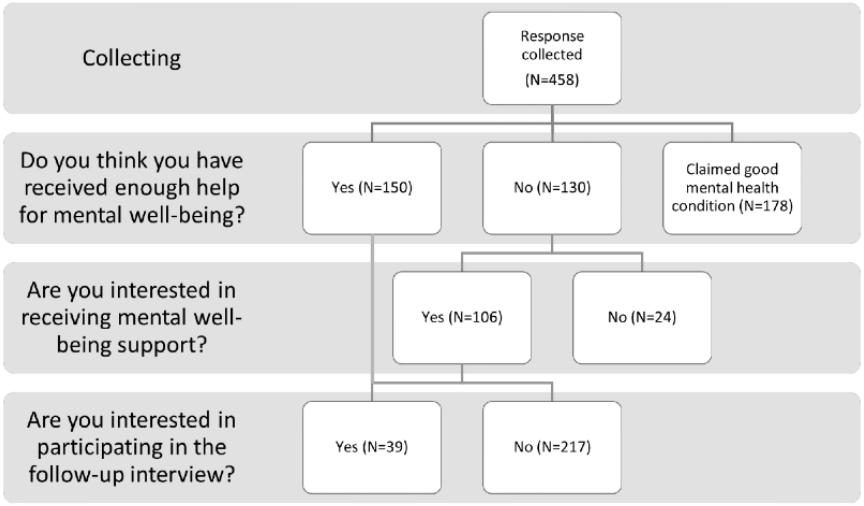
Recruitment

Participants in study 1 who expressed an interest in mobile-based mental health support technology and left their personal contact details for the follow-up interview, were invited to study 2. The logic of the selection is demonstrated in Figure 1.

A total of 39 eligible participants voluntarily left their contact details for the interviews in study 2. The final study size was 14, as interviewees started to discuss recurrent themes.

In addition, 5 managers from the same companies and organizations were invited to participate in the interviews.

Figure 1. Interview recruitment logic.



Interview Structure

In the study 2, detailed questions about perspectives on mental well-being were asked in the form of semistructured interviews held in Mandarin. The interview structure was designed to inspire the participants to openly share their opinions and ideas about digital support for workplace mental health.

Each interview took approximately 30 minutes, covering the following topics:

1. Experience of workplace stress and stigma toward mental health problems
2. Current way of dealing with workplace stress
3. Perspectives on mobile mental health support technologies in general
4. Feature requirements for a mobile mental health support technology
5. Other suggestions on mobile mental health support technology (art style, platform, privacy, etc)

Separate interviews were carried out with managers of each organization, discussing the current mental health support provided by the organization and their opinions on the mental health problems of employees in the workplace.

Data Analysis

Qualitative text data were manually transcribed from interviews by a Mandarin-speaking native Chinese researcher. Inductive thematic analysis was conducted on the transcribed text by 2 bilingual researchers (both had received university-level education in the United Kingdom) using a qualitative data analysis software, NVivo (QSR International). This study aimed to understand what motivates employees to use digital mental health support; therefore, the results were categorized into the following 3 main topics: their current ways of handling their stress, their opinions toward mental health in the workplace, and their ideas for digital mental health support technology. The text quotes were initially identified and differentiated into the above categories in the first round of coding and then iteratively subjected to inductive thematic analysis to develop 5 sets of views on stress, relaxation, workplace support, workplace mental health, and digital support technology. The last set of views about the technology consists of their requirements and aversions to features, media, structure, privacy, and esthetics.

Patient and Public Involvement

In this study, we did not inquire about diagnosis or treatment and did not talk about “patients” but rather about end users or the “public.” They were welcomed to discuss any interests and concerns regarding digital mental health support technologies in the semistructured interviews. End users were not involved in the study design.

Ethics Approval

This research was approved by the Imperial College Research Ethics Committee (ethics code 21IC6579).

Results

Survey Results

Demographics

A total of 458 participants completed the survey between June 2021 and November 2021 with a mean age of 32.7 (SD 9.5) years, ranging from 18-65 years (248/458, 54.3% female and 207/458, 45.3% male, as shown in Figures 2 and 3). Of the 458 participants, 14% (n=64) of the participants came from a manufacturing factory (RR [response rate]≈12.8%), 7.6% (n=35) of the participants from a software company (RR≈17.5%), 9.8% (n=45) of the participants from a private medical group (RR≈11.3%), 64% (n=293) of the participants from the government administration (RR≈20.9%), and 4.6% (n=21) from a university (8 students and others were teachers or administrative staff). Of all the participants, 2.4% (11/458) of the participants were undergoing clinical treatment for mental health problems (in any form, including medication and psychotherapy), and another 4.1% (19/458) of the participants had been treated before, with anxiety (18/30, 60%) and depression (16/30, 53%) being the 2 main conditions.

Of the 64 employees in the manufacturing company, which was located in a tier-3 city, 30 (47%) of employees were blue-collar production workers and 3 (5%) participants were either under treatment or had been previously treated. In tier-1 cities, this rate was higher (27/394, 6.85%), probably owing to the wider availability of mental health support resources.

Figure 2. Age distribution (in years).

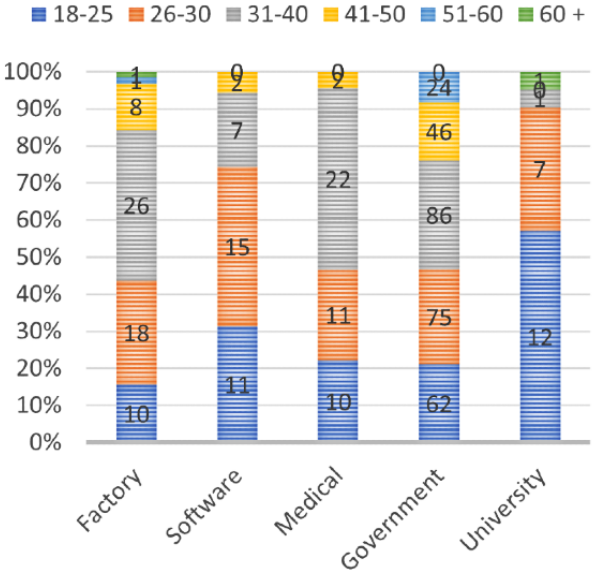
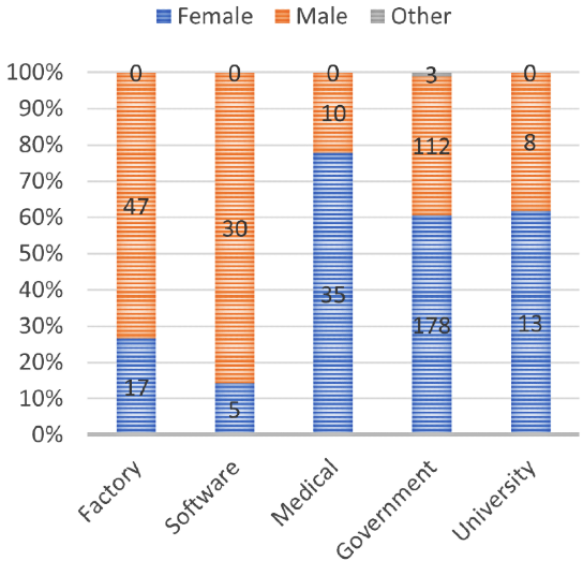


Figure 3. Gender distribution.



Smartphone Use

Understanding how the target population uses mobile phone technology provides valuable design insights. Of the 458 total participants, 457 (99.8%) participants reported smartphone use, with a mean use of 3.6 (SD 3.06) hours per day. Bandwidth is potentially an issue for 15% (69/457) of users; 63% (288/457) of the smartphone users have monthly data plans; and another 22.3% (102/457) have unlimited data plans. Most already see their phones as devices that can contribute to well-being; 73.5% (336/457) of smartphone owners use them for relaxation in their daily lives, with video and streaming apps being the top choice for mobile-based relaxation (209/457, 45.7%) followed by music apps (164/457, 35.9%), shopping apps (151/457, 33%), news and article apps (135/457, 29.5%), social media apps (133/457, 29.1%), and casual games (124/457, 27.1%). However, only 5 (1.1%) of the 458 participants used dedicated mental health support apps for relaxation.

Smartphone users were asked whether they would consider using a phone app to help them optimize their mental well-being. Furthermore, 50.5% (231/457) of the participants claimed that they were not interested in any form of support, as they were either satisfied with their current mental health status or had received sufficient support.

Additionally, 49.5% (226/457) of the participants demanded mental health support, 25.8% (118/457) would consider using an app, and another 21%(96/457) would consider taking help if it was in the form of a WeChat mini-program (a popular lightweight microapp embedded into the social messaging platform of WeChat) instead of a stand-alone app. Only 2.6% (12/457) of the participants responded negatively to help from digital technology. Of the 12 participants who resisted digital support, 10 (83%) of them had not been diagnosed with mental health difficulties and had not received any diagnosis of any kind. The other 2 (17%) employees who were receiving treatment preferred face-to-face clinical support from experts.

Current Mental Health Support

The participants' attitudes toward mental health support were investigated. Of the 458 participants, 38.9% (n=178) of the participants believed that they did not need any form of support, and another 32.8% (n=150) believed they had received sufficient help (note that this result is not the same as in section 3.1.2, as participants who are satisfied with their current support may still be interested in testing out such an app). For the 130 (28.4%) out of 458 participants who did not receive enough help, "I do not have time" (62/130, 47.7%), "I do not know where or how to seek help" (59/130, 45.4%), and "I do not want to worry friends and family" (54/130, 41.5%) were the top 3 concerns. Of the 130 participants, 106 (81.5%) wanted further mental health support.

Self-help (166/257, 64.6%) was the most desired way of support, followed by peer support (153/257, 59.5%) and face-to-face expert support (107/257, 41.6%). The preference for self-care is in line with acceptance of mobile-based mental health support.

Basic Psychological Needs at Workplace

According to the self-determination theory [25], the satisfaction of the 3 basic psychological needs of autonomy, competence, and relatedness has a positive relationship with psychological well-being and workplace engagement and is negatively related to ill-being. In this part of the survey, the participants were invited to participate in the BPNWS [23], which is a 12-item instrument on a 6-point scale. In the study by Brien at al [23], the French and English versions of the BPNWS were validated. The scale was translated into Chinese by 2 bilingual researchers. The results suggested that the reliability of the Chinese version of the scale was above good, with Cronbach α coefficients of 0.82, 0.92, and 0.93 for autonomous, competence, and relatedness, respectively. Compared with the French and Canadian teachers who were assessed using BPNWS by Brien et al [23], Chinese employees in general showed lower satisfaction scores for autonomy (mean 4.21, SD 1.40) but similar results for competence (mean 4.88, SD 1.06) and relatedness (mean 4.49, SD 1.16).

We note that the interpretation of the results is limited, given that the study was on teachers, whereas our study had participants from several industries.

ANOVA was conducted across the 5 industries separately for autonomy, competence, and relatedness. The mean scores for satisfaction with the 3 basic psychological needs are listed in (Table 1). No significant differences were observed in competence (P=.75) and relatedness (P=.25), but significant associations were shown in the mean scores for autonomy (P=.005) by industry. Post hoc tests suggested that the mean autonomy score of employees from the private medical company and university was significantly higher than the mean scores of employees from the government (P=.02 [medical company vs government] and P=.002 [university vs government]) and the factory (P=.01 [medical company vs factory] and P=.005 [university vs factory]). This is in line with the results of section 3.2.2, where in the interview, boredom was repeatedly mentioned by the government officers as a workplace stress source. Research in medical companies and universities could give employees more freedom to control their own work.

In addition to understanding what might drive mental ill health, it is useful to understand what drives the individuals to seek help.

Table 1. Mean scores (and SD) of basic psychological needs satisfaction for the 5 industries.

	Software company	Medical company	University department	Government office	Factory
Autonomy, mean (SD)	4.29 (1.02)	4.64 (0.81)	4.68 (0.69)	4.10 (1.13)	4.18 (0.82)
Competence, mean (SD)	4.81 (1.15)	5.02 (0.61)	4.93 (0.60)	4.87 (0.98)	4.89 (0.86)
Relatedness, mean (SD)	4.49 (1.25)	4.54 (0.86)	4.48 (0.68)	4.54 (1.08)	4.21 (1.07)

Feature Preferences

The 214 participants who showed an interest in the smartphone-based mental health app were asked to select their desired features. There were no limitations to the number of features selected. The top 10 feature preferences are summarized in Table 2, showing the following priorities.

Table 2. Feature preferences (N=214).

Rank	Features	Participants, n (%)
1	Mood-fix tool	127 (59.3)
2	Talk with experts	121 (56.5)
3	Book and article recommendation	117 (54.7)
4	Self-assessment quiz with feedback	109 (50.9)
5	Ability to record the mood triggers	109 (50.9)
6	Dashboard showing progress	108 (50.5)
7	Emotion tracking	101 (47.2)
8	Workouts and exercises	100 (46.7)
9	Suggestions based on self-assessments	98 (45.8)
10	Relaxing music	98 (45.8)

Interview Results

Stress Source

Participants reported 9 main sources of workplace-related stress. These stress sources are classified into 2 broad categories based on the Job Demands-Resources model [26]. An imbalance between demands on employees and the resources they possess to deal with those demands leads to workplace stress.

Job demands including dealing with unreasonable clients, meeting deadlines, giving a public speech, reaching target key performance indicators, and taking on responsibilities are causing job strain. The lack of job resources had a further negative impact, including boredom, managerial difficulties, insufficient support from teammates, and low confidence in obtaining promotions. Both high job demands and limited job resources contributed to stressful working experiences.

In addition to work-related issues, interviewees expanded the topic to express their concerns for other stress sources, including family, finance, commuting, and health conditions. For future development of workplace mental health support, we need to bear in mind that we should not only solely focus on work-related problems but also pay equal attention to other factors.

Methods to Relieve Stress

Talking about ways to relieve stress, participants held a common ground that “To get to the root of the stress problem, the best solution is to address the issue as quickly as possible.” However, if it could not be solved in a timely manner or required support from others like experts, it was common for the employees to try to address it by talking to their friends and colleagues, practicing outdoor activities, or relying on digital devices for relaxation (see section 3.1.2).

Employees also mentioned that their companies or organizations provided human resource consulting, team building activities, mental health training lessons, relaxation rooms, gyms, and massage sofas as mental health support.

Attitudes Toward Mental Health

Most employees (who were coded as [E1] to [E14] to label their quotes) acknowledged the importance of mental health education and support when discussing their opinions on mental health:

> *It is common for people to feel depressed or anxious at times. Everyone has a certain threshold for taking up stress. When that threshold is reached, he will lose his control for a while.* [E5]

Most interviewees showed understanding and sympathy toward people with mental health difficulties. They could recognize depression or anxiety as real medical illnesses and would actively encourage those experiencing these problems to speak out to them or seek professional support.

However, when they evaluated people with mental health difficulties from the perspective of working partners, interviewees would lower their expectations of their colleagues’ abilities in many aspects including communication, teamwork, capability, problem solving, time management, and work motivation.

Interviewees showed a tendency to maintain a distance from mentally unwell people for their own safety and that of the person:

> *People who are affected by negative emotions are prone to anger and temper tantrums.* [E2]

> *If I knew this person had anxiety, I would definitely be wary of this person. Because I would have concerns about that if he suddenly became irritable.* [E7]

Stigma was another topic that was repeatedly discussed in the interviews. Interviewees described most people as holding “conservative” views on mental health, as discrimination toward mental illness is common in the Chinese culture. In Chinese, the word “mentally ill” (“精神病”) is widely used as a cursing word to criticize the insanity behind someone’s misbehavior. Advising someone to go see a psychologist is considered “extremely impolite and unacceptable.” Thus, employees with mental difficulties will make their best efforts to hide their conditions to avoid underestimation of working abilities by their colleagues or leaders. However, based on their own experiences

and examples around them, most (10/14, 71.4%) interviewees admitted that covering it up was merely a means of protecting the sufferer from further social pressure, which might lead to a "vicious circle" of stress accumulation. This creates the requirement for self-help technology for mental health support.

Company Organization Support

Government administration interviewees claimed that they were not provided with mental health support services. However, there was a "relax room" in every office site where employees could rest, sit on the massage sofa for a while, and play table tennis. In addition, once a year, a lecturer might be invited to talk to staff about mental training. However, staff found it less helpful because of the few training programs.

Employees in manufacturing and software companies referred to talking with human resource managers and team building as mental well-being support provided by the company.

Managers in these companies claimed that the reasons they did not set up professional mental health support included the cost to the company and the privacy concerns of employees that would hinder their engagement. It was also difficult for them to assess the benefits of investing resources into mental health support.

Features Requirements

From the qualitative analysis of the 14 interviews, 8 topics were repeatedly identified for digital mental health support. These features include the following:

1. Self-assessment
2. Education
3. Mood-fix tools
4. Self-help exercise
5. Seek help from experts or community
6. Design style
7. Worries
8. Trust

We will elaborate more on each topic.

Self-assessment

The first common expectation of employees was to judge their current state of mental health:

> *I would like the app to give me an understanding of my mental wellbeing conditions, preferably through a structured medically supported diagnosis. There should be questions to determine what level of problem you are at, thus suggest how to relax, whether you need to go to hospital, or even, whether you need to be hospitalised.* [E1]

> *Because I don't have psychological knowledge, I don't know if I have a problem. I would like to judge it for myself with the help of theoretical instructions.* [E4]

Education

Employees showed interests in learning trustworthy psychological knowledges:

> *I am looking for books and articles on the subject. I want to logically sort out where my emotions come from.* [E11]

> *It is like getting a cold, once a man sneezes, his common senses tell him that he is having a cold. One should learn some common senses while he is still healthy. Thus, once he does get into unpleasant situation, he knows he may be depressed or not in a normal state of mind.* [E5]

Participants also expected that with the help of the self-assessment feature, it would be most helpful to recommend personalized psychoeducation contents to users.

In addition, employees were keen to acquire workplace-specific knowledge as part of their mental self-help. This included how to overcome workplace crises, how to communicate with leaders and colleagues, how to deal with workplace bullying, and how to handle stress through time management:

> *When you encounter a crisis at work, you may need some tips from psychological perspectives. There are so many contradicting articles on the internet, but I don't know what is right and what is wrong, so I hope I can get access to some articles or content that is really valuable.* [E11]

> *It will be interesting to learn about how other countries have dealt with workplace mental health. Because China is at a very early stage in this area, people may not have any idea, so it helps to give some examples of how foreign countries with a more developed mental health support system deal with these problems.* [E6]

Short videos are the most welcomed media for receiving psychoeducation, and many participants mentioned TikTok as their favorite way of relaxation using digital devices.

Mood-Fix Tools

Participants expressed a strong demand for tools that could help them quickly fix their mood during a workplace mental crisis that leads to moments of depression or anxiety.

Based on their current methods of quick mood-fixes using digital devices, they suggested soothing music, humorous short stories, casual games, and breathing exercises.

Self-help Exercises

Employees demanded mental health supportive exercises, both spiritual and physical.

Although meditation is less known as an evidence-based practice for mental health support, its connection to Buddhism draws interest, as Buddhism plays a vital role in the spiritual aspects of Chinese culture:

> *There are some useful practices in Buddhism, such as meditation, selflessness, letting go of obsessions, yoga. It's about de-emphasising self-consciousness and relationships with people, making yourself pure and calm. Relieving your stress by thinking of problems in a more objective perspective.* [E6]

> *I want to learn meditation through a tutorial with daily exercises.* [E4]

However, some interviewees raised concerns over conditions for meditation:

> *I think meditation is very useful, but it requires willpower and a quiet and relaxing environment, which many people don't have.* [E11]

Physical activities including exercises and sleep quality improvement exercises are also demanded:

> *Add some exercises, such as simple stretches or office exercises where you can practice in a chair.* [E6]
>
> *Sometimes when we go to bed and wake up, many of our worries are gone, but if we don't get a good night's sleep, even if we don't have any worries, they will arise. The endocrine system is also affected. It would be good if there was an app that could help with sleep.* [E12]

Seek Help From Experts or Community

Some employees expressed a desire to communicate with experts or peers. However, others had different attitudes toward price, efficacy, and privacy:

> *If they can't solve it by self-help, the app can recommend some psychologists or other clinical treatments.* [E13]
>
> *I think that, for social purposes, it might be more effective to find people on the Internet who share the same goals and talk about them. But there are ethical issues involved.* [E4]
>
> *Responsible people may be reluctant to talk to their families about the pressures at work. Some incognito channels of communication with the leader can be helpful.* [E10]
>
> *I often read psychological articles in Zhihu (A Chinese question-and-answer website). Sometimes I don't like things that are particularly official. It's interesting to see the answers to a certain question from various perspectives. People want to find others who can understand them. However, in a community like this, where you ask a question and others will answer and discuss it, sometimes it can distort your perceptions and make the problem even worse. Sometimes, you still need to take guidance from professionals.* [E11]
>
> *I think that when it comes to approaching a psychiatrist, the first concern is probably the price. You will be worried what if, in fact, these people are merely there to give you ambiguous advice for money.* [E6]

Design Styles

Participants actively described their preferences for digital support in the study, and their ideas were summarized using the following 3 keywords as shown in Textbox 1.

Textbox 1. Keywords and ideas.

Mind-opening

- *"contents that exposes me to new stuff and holds my attention"* [E3]

Straightforward

- *"The only thing I am willing to do when I am stressful is to do simple relaxing activities following instructions from others."* [E12]
- Rather than making the users think overly complicated, the participants preferred plain and straightforward wording and exercises.

Flexible

- *"I tend to have more freedom. Don't want to stick to all those rules and regulations. I prefer a bit of simplicity."* [E14]

Worries

Of the 14 participants, 5 (36%) questioned the possible effectiveness of using a mental health support technology. For example:

> *Being stressful is because the task is not finished and there are many difficulties in finishing it. If you let me play games or listen to music for a short time, because I still have unfinished business in my mind, in the long term it won't help and may put on further stress.* [E2]
>
> *When I feel stressful, temporarily letting go of my anxiety could make me even more stressful afterwards.* [E11]

Employees were also concerned about receiving contents that they found inappropriate, especially tedious sermons and advertisements:

> *I hate being lectured, 'You should do this,' 'You shouldn't do that.' I'd rather have no one around than someone there to lecture me.* [E10]
>
> *When you are stressed, you don't want to study, you don't want to get overwhelmed by lecturing texts, and you certainly don't want to see any advertisements."* [E13]
>
> *I don't want to see any advertisement. I will doubt if the contents are really valid or are they just trying to lure me to pay for their services.* [E3]

Trust

At the end of the study, participants were asked about their trust in provider of mental health support technology.

All the interviewees showed a lack of trust in their companies, fearing that their privacy would be violated by exposing their mental health conditions to their employers. In contrast, people place more trust in government and nonprofit organizations.

Managers' Perspectives

In addition to the interviews with employees, 5 managers (1 from the government office, 3 from the factory, and 1 from the software company) were invited for interviews before starting the survey. They were either human resource directors or executives of the companies or organizations. Team key performance indicators and team management were the primary sources of stress for managers in the workplace.

From a personal perspective, most managers claimed that employees with mental health problems were understandable and should not be discriminated against. However, from the standpoint of the company or organization, they were not likely to hire those with mental health issues. This sentiment is reflected in a quote from a manager of a manufacturing company:

> *Workers live and eat together so that it will have a bad influence.*
>
> *Most of the time, I will not hire someone with mental health or anxiety problems. Unless his or her mental health condition somehow suits his position. For example, if his or her anxiety is caused by perfectionism, he or she may be a suitable candidate for quality inspection tasks.*
>
> *I will not hire someone with mental health problems. Especially for salespeople. I don't believe they have the ability to communicate well with the customers.*

When managers were asked about the measures they would consider taking if one of the employees was found to experience mental problems, they held different opinions:

> *Employees with anxiety problems are acceptable if they do not negatively affect others. A negative employee may have a bad influence on the work atmosphere, and thus make others lose confidence in dealing with the current problem and hard to collaborate.*
>
> *Anxiety may be caused by incompetence. However, most of the time, the problem is that he or she is put into the wrong position. Changing his type of work may help.*
>
> *I will encourage them to stop working and seek interventions. Employees with severe mental health problems may pose physical threats to their colleagues.*
>
> *If we found some employee currently suffering from anxiety problems, we will advise him to quit the job.*

The managers mentioned that their expectations for employees with mental problems would be lowered, especially in their work performance and time management ability.

All the managers believed that a mental health support app would benefit both the employees and the company. However, cost is the top concern that prevents companies from setting up mental health services for employees. Moreover, even if managers are willing to provide the services, they do not know where to start. This is because of the lack of professional training and guidance in relevant fields for managers.

Discussion

Principal Findings

The results of study 1 showed that 46.3% (106/226) of the participants who were willing to receive mental health support felt that they had not received enough help. Participants in study 2 also reported a limited provision of mental health support services by the companies. This finding is in line with previous studies (as described in section 1), which reiterates that stigma and lack of resources deter people in China from seeking mental health support.

The results from section 3.1.4 suggest that Chinese employees may feel less autonomous than their French and Canadian peers, with mean satisfaction scores lower than those of both Western countries. However, this result requires further validation, as the satisfaction of autonomy needs differs across industries in China. The evidence suggests that support could be developed to prevent stress by guiding employees to satisfy their sense of autonomy, as the lack of autonomy can lead to workplace stress according to the theory of Job Demands-Resources model [26].

Although the participants in study 2 showed empathy and understanding toward people with mental health conditions, discrimination still widely exists. People with mental health conditions are criticized for their weaknesses and tend to underestimate the various abilities required by their jobs. Employees are usually rejected or even dismissed if their mental health conditions are exposed to the company. Thus, they tend to cover their true emotions, leaving them less likely to receive appropriate help and support. This self-concealing of emotions and feelings together with the Asian culture of "not wanting to worry friends and family" makes self-help (166/257, 64.6%) the most desired way of digital mental health support.

As employees spend a lot of time on their smartphones with internet access, it is common to rely on them for relaxation and mental health support. Although very few of them benefited from dedicated mental health support apps, most were willing to try out such a technology, provided it was effective.

As explained in section 3.1.5, study 1 showed employees' requirements for features in the 5 categories: self-assessment, self-guided support, psychoeducation, emotion tracking, and expert counseling.

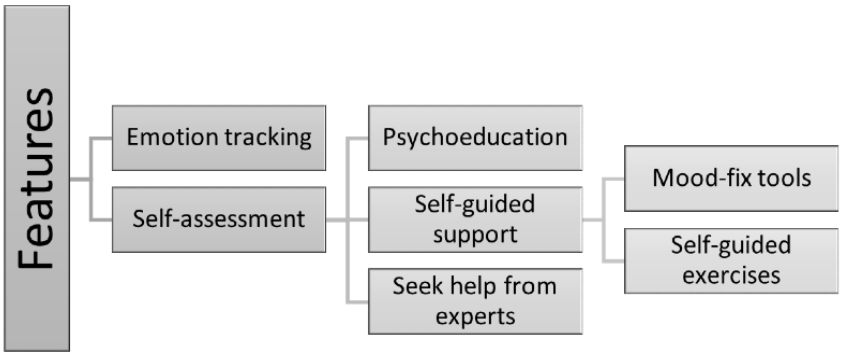
Combined with the results of study 2, the features are summarized in the logic shown in (Figure 4). In addition to the feature of emotion tracking and corresponding triggers, employees would like to take self-assessments to understand

their current mental health conditions and based on this receive psychoeducational material, access to self-guided support, and seek help from experts accordingly.

The statistics also favored the deployment of digital mental health support on a WeChat "mini-program," as users can access the contents in the mini-programs immediately without downloading, installing, or signing up.

As revealing one's mental health problems can lead to underestimation and alienation, people take their privacy seriously. As discussed in section 3.2.5, employees are more likely to trust and adopt support from publishers such as the government or nonprofit organizations that are not related to their employment. As a part of the "National Mental Health Working Plan (2015-2020)," the Chinese government is actively experimenting community-based support to promote mental health [11]. With central government endorsement, introducing self-help mental health support by the community is more likely to result in increased motivation than that by the company. This could impact the development of Employment Assistance Program services in China. The providers need to convince employees that their information and records will not be accessible to their employers, gain their trust, and promote their motivation to use the service.

Figure 4. Digital support features requirements.



Limitations

The main limitation of this study is the potential bias in the sample participants. Although the participants were selected from 5 companies or organizations with diverse backgrounds, all these companies and organizations are located in large cities with more than 4 million residents. Except for 6.6% (30/458) of the factory workers, all other employees in the study could be classified as white-collar workers living in urban areas. The mental health condition and support preferences of workers living in less developed cities and rural areas could be quite different from these results [4,5].

The limitation of the survey dissemination method is that if the employee was not on the web portal during the spread of the survey link, the group chat was likely to be flooded by other messages, so the employee might not have the chance to read the invitation at all. Therefore, it was impossible to calculate the number of employees who read the message, resulting in a much lower RR.

In addition, the results are likely to have been influenced by self-selection bias, especially in study 2 where voluntary participants are likely to represent individuals who are more open to mental health issues.

Future Work

Design engineers, working in collaboration with mental health researchers, can contribute to understanding the issues that engage or disengage users with digital mental health interventions. In the West, such collaborations are increasingly common as shown by the emergence of journals and workshops such as [27,28]. In China, the adoption of such design procedures is still in the exploration stage. We aim to bridge this gap by designing, developing, and piloting further research.

Conclusions

Our study across 5 companies and organizations in China showed that the proportion (150/458, 32.8%) of employees in these companies and organizations who were accessing any form of mental health support was very low, both from clinical treatment and digital support via smartphones. However, more than a quarter (118/457, 25.8%) of the respondents reported that they liked much more smartphone-based mental health support, indicating an exceptionally large need and demand for such services. The qualitative interviews demonstrated problems caused by stigma around mental ill-being, such as unwillingness to disclose and low help-seeking rates, either in clinical services or within the company or organization. This suggests that the way forward is to develop confidential and anonymized apps that enable people to undertake self-help to improve their mental health. Certain elements of features were deemed very desirable, including tools to help people fix their moods and professional guidance either in person or with related books and articles. This will help us to think about how to develop the content of a self-help app to enable Chinese employees to improve their mental health not only focusing on work stress but also on other stressors in their lives.

Acknowledgments

SS (lead author) conceived the study, and it was further refined by RAC and CM. SS developed the detailed methodology and study design and applied it for ethical approval. The recruitment of participants was led by SS with the help of MT. Data analysis was conducted jointly by SS and ZZ. Discrepancies between SS and ZZ regarding thematic analysis were resolved by RAC and CM. All the authors (SS, ZZ, MT, CM, NG, and RAC) reviewed, commented, and edited the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Riba MB, Parikh SV, Greden JF. Mental Health in the Workplace: Strategies and Tools to Optimize Outcomes. Cham, Switzerland: Springer; 2019.
2. Farmer P, Stevenson D. Thriving at Work: a review of mental health and employers. Department for Work and Pensions and Department of Health and Social Care, United Kingdom Government. 2017 Oct 26. URL: https://www.gov.uk/government/publications/thriving-at-work-a-review-of-mental-health-and-employers [accessed 2021-05-14]
3. 2020 China workplace employees' health risk report. Mercer Marsh Benefits. 2021. URL: https://www.mercer.com.cn/our-thinking/health/2020-China-healthiest-workplace-employee-survey.html [accessed 2022-12-15]
4. Li X, Zhang W, Lin Y, Zhang X, Qu Z, Wang X, et al. Pathways to psychiatric care of patients from rural regions: a general-hospital-based study. Int J Soc Psychiatry 2014 May;60(3):280-289. [doi: 10.1177/0020764013485364] [Medline: 23704113]
5. Zhang W, Li X, Lin Y, Zhang X, Qu Z, Wang X, et al. Pathways to psychiatric care in urban north China: a general hospital based study. Int J Ment Health Syst 2013 Sep 10;7(1):22 [FREE Full text] [doi: 10.1186/1752-4458-7-22] [Medline: 24020825]
6. Fang M, Hu SX, Hall BJ. A mental health workforce crisis in China: a pre-existing treatment gap coping with the COVID-19 pandemic challenges. Asian J Psychiatr 2020 Dec;54:102265 [FREE Full text] [doi: 10.1016/j.ajp.2020.102265] [Medline: 32622031]
7. Farndon H. HCPC Registered Psychologists in the UK. Leicester, UK: British Psychological Society; 2016.
8. World Health Organization. Mental Health ATLAS 2017. Geneva, Switzerland: World Health Organisation; 2018.
9. Xu J, Wang J, Wimo A, Qiu C. The economic burden of mental disorders in China, 2005-2013: implications for health policy. BMC Psychiatry 2016 May 11;16:137 [FREE Full text] [doi: 10.1186/s12888-016-0839-0] [Medline: 27169936]
10. Mental health: strengthening our response. World Health Organization. 2022 Jun 17. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2022-01-13]
11. Liang D, Mays VM, Hwang WC. Integrated mental health services in China: challenges and planning for the future. Health Policy Plan 2018 Jan 01;33(1):107-122 [FREE Full text] [doi: 10.1093/heapol/czx137] [Medline: 29040516]
12. Park L, Xiao Z, Park JM. Mental health care in China: recent changes and future challenges. Harvard Health Policy Rev 2004 Nov;6(2):35-45.
13. Lu L, Kao SF, Siu OL, Lu CQ. Work stress, Chinese work values, and work well-being in the Greater China. J Soc Psychol 2011;151(6):767-783. [doi: 10.1080/00224545.2010.538760] [Medline: 22208113]
14. Montenegro CR, Ortega F. Thinking beyond implementation: context and culture in global mental health. BMJ Glob Health 2020 Dec;5(12):e004539 [FREE Full text] [doi: 10.1136/bmjgh-2020-004539] [Medline: 33380416]
15. Scheutzow J, Attoe C, Harwood J. Acceptability of web-based mental health interventions in the workplace: systematic review. JMIR Ment Health 2022 May 11;9(5):e34655 [FREE Full text] [doi: 10.2196/34655] [Medline: 35544305]
16. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017 Dec 21;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
17. Stratton E, Jones N, Peters SE, Torous J, Glozier N. Digital mHealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021 Aug 01;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
18. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 01;45(6):560-576 [FREE Full text] [doi: 10.5271/sjweh.3839] [Medline: 31184758]
19. The 47th statistical report on China's Internet development. China Internet Network Information Center. 2021 Feb. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202104/P020210420557302172744.pdf [accessed 2021-03-12]
20. Mak WW, Chan AT, Cheung EY, Lin CL, Ngai KC. Enhancing Web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]

21. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR Mhealth Uhealth 2019 Nov 07;7(11):e13236 [FREE Full text] [doi: 10.2196/13236] [Medline: 31697245]
22. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
23. Brien M, Forest J, Mageau GA, Boudrias JS, Desrumaux P, Brunet L, et al. The basic psychological needs at work scale: measurement invariance between Canada and France. Appl Psychol Health Well Being 2012 Jul;4(2):167-187. [doi: 10.1111/j.1758-0854.2012.01067.x] [Medline: 26286976]
24. Peters D, Deady M, Glozier N, Harvey S, Calvo RA. Worker preferences for a mental health app within male-dominated industries: participatory study. JMIR Ment Health 2018 Apr 25;5(2):e30 [FREE Full text] [doi: 10.2196/mental.8999] [Medline: 29695371]
25. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY, USA: Guilford Press; 2018.
26. Bakker AB, Demerouti E. The Job Demands ‐ Resources model: state of the art. J Manag Psych 2007 Apr 03;22(3):309-328. [doi: 10.1108/02683940710733115]
27. Calvo RA, Dinakar K, Picard R, Maes P. Computing in mental health. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 May Presented at: CHI EA '16; May 7-12, 2016; San Jose, CA, USA p. 3438-3445. [doi: 10.1145/2851581.2856463]
28. Calvo RA, Dinakar K, Picard R, Christensen H, Torous J. Toward impactful collaborations on computing and mental health. J Med Internet Res 2018 Mar 09;20(2):e49 [FREE Full text] [doi: 10.2196/jmir.9021] [Medline: 29426812]

Abbreviations

BPNWS: Basic Psychological Needs at Work Scale
MARS: Mobile App Rating Scale
RR: response rate

Edited by A Kushniruk; submitted 09.07.22; peer-reviewed by J Scheutzow, CY Lin; comments to author 04.10.22; revised version received 27.10.22; accepted 09.11.22; published 22.12.22.

Please cite as:
Sun S, Zhang Z, Tian M, Mougenot C, Glozier N, Calvo RA
Preferences for a Mental Health Support Technology Among Chinese Employees: Mixed Methods Approach
JMIR Hum Factors 2022;9(4):e40933
URL: https://humanfactors.jmir.org/2022/4/e40933
doi:10.2196/40933
PMID:36548027

Original Paper

Desired Characteristics of a Clinical Decision Support System for Early Sepsis Recognition: Interview Study Among Hospital-Based Clinicians

Jasmine A Silvestri[1*], MPH; Tyler E Kmiec[1*], MPH; Nicholas S Bishop[1], BA; Susan H Regli[2], PhD; Gary E Weissman[1,3,4,5], MD, MSHP

[1]Palliative and Advanced Illness Research Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States
[2]University of Pennsylvania Health System, Philadelphia, PA, United States
[3]Division of Pulmonary, Allergy and Critical Care Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States
[4]Leonard Davis Institute of Health Economics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States
[5]Penn Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States
[*]these authors contributed equally

Corresponding Author:
Gary E Weissman, MD, MSHP
Palliative and Advanced Illness Research Center
University of Pennsylvania Perelman School of Medicine
300 Blockley Hall
423 Guardian Drive
Philadelphia, PA, 19104
United States
Phone: 1 215 746 2887
Email: gary.weissman@pennmedicine.upenn.edu

Abstract

Background: Sepsis is a major burden for health care systems in the United States, with over 750,000 cases annually and a total cost of approximately US $20 billion. The hallmark of sepsis treatment is early and appropriate initiation of antibiotic therapy. Although sepsis clinical decision support (CDS) systems can provide clinicians with early predictions of suspected sepsis or imminent clinical decline, such systems have not reliably demonstrated improvements in clinical outcomes or care processes. Growing evidence suggests that the challenges of integrating sepsis CDS systems into clinical workflows, gaining the trust of clinicians, and making sepsis CDS systems clinically relevant at the bedside are all obstacles to successful deployment. However, there are significant knowledge gaps regarding the achievement of these implementation and deployment goals.

Objective: We aimed to identify perceptions of predictive information in sepsis CDS systems based on clinicians' past experiences, explore clinicians' perceptions of a hypothetical sepsis CDS system, and identify the characteristics of a CDS system that would be helpful in promoting timely recognition and management of suspected sepsis in a multidisciplinary, team-based clinical setting.

Methods: We conducted semistructured interviews with practicing bedside nurses, advanced practice providers, and physicians at a large academic medical center between September 2020 and March 2021. We used modified human factor methods (contextual interview and cognitive walkthrough performed over video calls because of the COVID-19 pandemic) and conducted a thematic analysis using an abductive approach for coding to identify important patterns and concepts in the interview transcripts.

Results: We interviewed 6 bedside nurses and 9 clinicians responsible for ordering antibiotics (advanced practice providers or physicians) who had a median of 4 (IQR 4-6.5) years of experience working in an inpatient setting. We then synthesized critical content from the thematic analysis of the data into four domains: clinician perceptions of prediction models and alerts; previous experiences of clinician encounters with predictive information and risk scores; desired characteristics of a CDS system build, including predictions, supporting information, and delivery methods for a potential alert; and the clinical relevance and potential utility of a CDS system. These 4 domains were strongly linked to clinicians' perceptions of the likelihood of adoption and the impact on clinical workflows when diagnosing and managing patients with suspected sepsis. Ultimately, clinicians desired a trusted and actionable CDS system to improve sepsis care.

Conclusions: Building a trusted and actionable sepsis CDS alert is paramount to achieving acceptability and use among clinicians. These findings can inform the development, implementation, and deployment strategies for CDS systems that support the early detection and treatment of sepsis. This study also highlights several key opportunities when eliciting clinician input before the development and deployment of prediction models.

(JMIR Hum Factors 2022;9(4):e36976) doi:10.2196/36976

KEYWORDS

sepsis; predictive information; clinical decision support; human factors; sepsis onset

Introduction

Background

Sepsis, a life-threatening dysregulation of the immune system in response to an infection, is a significant risk for patients and a major burden for health care systems in the United States, with over 750,000 cases annually and total costs nearing US $20 billion [1,2]. The hallmark of effective sepsis treatment is early recognition and initiation of broad-spectrum antibiotic therapy [3,4]. However, sepsis is characterized by high diagnostic and prognostic uncertainty, which often results in delayed recognition and treatment, especially among patients who develop sepsis in hospitals.

To facilitate the timely recognition and management of sepsis, several machine learning prediction algorithms have been developed and integrated into electronic health record (EHR)–based alerts and clinical decision support (CDS) systems [5-7]. Although deployment of such CDS systems is common, there is little high-quality evidence to suggest that they are reliably effective in improving care processes or clinical outcomes [5,6]. Several prior studies have identified barriers to successful integration of sepsis CDS systems into clinical practice, including poor diagnostic accuracy, poor education and implementation strategies, and clinician mistrust of an unfamiliar system [8,9]. At the same time, human factors research into sepsis-specific CDS systems has focused on the display of visual information but has overlooked team-level dynamics and clinician-level affective and cognitive influences [10-12]. User interfaces for predictive models, often built with complex statistical learning algorithms, may present clinicians with outputs that are difficult to explain and that sometimes contrast with clinical intuition, thereby decreasing the likelihood of adoption [12,13].

Objective

To address these research gaps, we sought to elicit perspectives on and preferences for a hypothetical, sepsis-focused predictive CDS system from a multidisciplinary group of hospital-based clinicians who regularly care for patients suspected of sepsis. We used qualitative semistructured interviews informed by human factors methodology to identify important clinician perspectives on the clinical and team-level context for using an alert. Our goal was to elicit this information to inform the design and implementation of a future early-warning sepsis CDS tool.

Methods

Overview

To investigate how a sepsis CDS system might integrate into clinical workflows and to elicit clinicians' perspectives on and preferences for prediction information, we used human factors methodologies of contextual inquiry and cognitive walkthrough (we met over video rather than physically in the participants' workspaces as a necessary modification because of constraints of the COVID-19 pandemic).

Vignette and Simulated Chart Development

We first engaged clinicians in diagnosis and decision-making through clinical vignettes in which sepsis was on the differential diagnosis. The study team prepared 2 vignettes (Multimedia Appendix 1) for each patient with possible sepsis varying in severity. Clinical decision-making regarding sepsis treatment is imbued with uncertainty related to both the diagnosis and the severity of the presentation [14]. To explore how a sepsis CDS system might optimally support such decision-making across the full range of diagnostic and prognostic uncertainty, we varied this uncertainty in 2 distinct vignettes. The first vignette was created as a straightforward case that met all sepsis criteria and had no obvious competing diagnosis. The second vignette was created to have higher diagnostic uncertainty, in which the patient met fewer formal criteria and had a broader differential diagnosis of potentially causative disease processes but presented with a higher severity of illness. Our team worked with an EPIC (Epic Systems Corporation) build specialist to develop 2 simulated patient charts in the EHR that reflected the clinical courses described in the vignettes. In addition to reviewing the written vignette, each participant in the first round of interviews was asked to review the simulated data in the EHR and to verbalize their thoughts and considerations in a simulated clinical evaluation. After the research team conducted a preliminary analysis, a list of factors pertaining to sepsis was developed and presented to round 2 participants for review in tandem with the vignettes. This 2-step process allowed the study team to member check our findings from previous interviewees and ensure that we captured a comprehensive list of the relevant factors assessed, completing a differential diagnosis for patients with possible or suspected sepsis.

Study Population and Recruitment

From September 2020 to February 2021, we recruited a convenience sample of 5 physicians, 4 advanced practice providers (APPs) (nurse practitioners and physician assistants), and 6 bedside nurses who cared for inpatients at the University of Pennsylvania Health System. During round 2 of interviews,

we determined that we had reached saturation after 3 participants did not contribute any additional factors to our comprehensive list of data points considered during a hypothetical differential diagnosis of vignette patients. We identified eligible participants through department leaders and staff lists, and sent email invitations to participate in the study. We used a 2-step hierarchical recruitment approach. First, we used purposive sampling to identify a range of specialty wards in the University of Pennsylvania Health System. Second, we recruited clinicians directly via email from these wards using convenience sampling. We chose this method to facilitate the inclusion of a range of perspectives from clinicians in different care settings while also completing recruitment in a timely manner. All participants verbally acknowledged their consent to participate at the onset of the web-based interview and received a US $50 gift card for completing the study.

Ethical Considerations

This protocol was deemed exempt by the institutional review board of the University of Pennsylvania (protocol number 843819).

Interview Guide, Data Collection, and Analysis

Interviews were conducted via videoconference by either a qualitative research specialist (JS) or a doctorate-level coinvestigator (SR) with extensive qualitative training. During the interviews, responses were noted by the clinical research coordinators of the research team (TK and NB). To reduce bias regarding the diagnosis of hypothetical patients, no clinically trained research staff or experts in sepsis were present during the interviews. As nurses and physicians discussed the desired aspects of sepsis alerts or other topics of interest, interviewers earnestly followed those lines of inquiry, asking open-ended questions to elicit comprehensive responses. For each interview, participants were presented with both vignettes (a straightforward vignette followed by a more complex and diagnostically ambiguous vignette) and asked questions about their communication and decision-making processes. A subsequent round of EHR walkthroughs and review of pertinent factors when considering sepsis followed. After mentally engaging each participant in this task, we asked questions about their prior experiences with and preferences for predictive information relevant to caring for patients suspected of sepsis. Each interview lasted approximately 45 minutes, was recorded and professionally transcribed, and was deidentified before analysis. This study reports the findings related to preferences for and perspectives on prediction information, while the results focused on decision-making will be reported separately.

We first interviewed 3 clinicians (1 physician and 2 bedside nurses) who were experienced in sepsis management to test our interview guide and simulated EHR data for accuracy and effectiveness. The interview guide was updated iteratively to reflect the emerging themes and questions. All subsequent interviews took place with clinicians who staffed ward units (general medicine, oncology, pulmonary, neurosurgery, and gerontology). The initial interview questions were designed to engage participants in active decision-making and to elicit a differential diagnosis. Subsequent questions sought to elicit the perceived potential impacts of a predictive sepsis alert on workflow and decision-making when diagnosing and managing a patient suspected of sepsis, including preferences for and acceptability of a potential future sepsis alert.

In this qualitative study, we used an abductive analytic approach in which the existing theory can be built upon a combination of inductive and deductive approaches to coding and emphasizing new or surprising findings [15,16]. To inform the development of our codebook, 3 members of the team (JS, SR, and GW) first independently identified themes and then met to discuss commonalities. Additionally, our interview guide was developed to include inquiry into several factors identified in previous research as integral to the development of CDS [17]. Half of the transcripts were reviewed and coded by at least two members of the research team (JS, SR, and GW); all disagreements were reconciled through consensus. The remaining transcripts were coded by a single member of the research team (JS).

Results

Overview

We conducted 15 interviews with 5 physicians, 4 APPs, and 6 nurses (Table 1). Through thematic analysis, we identified 4 broad themes linked to the likelihood of adoption and their impact on clinical workflows when diagnosing and managing patients suspected of sepsis. The first theme was clinician perceptions of prediction models, including both positive and negative sentiments that shaped how clinicians viewed predictive information. The second theme was previous experiences of clinician encounters with predictive information and risk scores, both in the context of the local health system and with nationally recognized tools for sepsis identification. The third theme centered on the desired characteristics of a CDS system build and included predictions, supporting information, and delivery methods for a potential alert. The fourth theme included the clinical relevance and potential utility of a CDS system for its intended audience. These themes, including codes, definitions, and examples of each major theme, are detailed in Table 2. In addition, select illustrative quotes are provided to provide additional context to the identified themes.

Table 1. Clinical cohort characteristics (N=15).

Participant characteristic	Value
Age (years), mean (SD)	
18-24	1 (6.7)
25-34	13 (86.7)
35-44	1 (6.7)
Sex, n (%)	
Male	4 (27)
Female	11 (73)
Race, n (%)	
Asian or Asian American	2 (13)
White	11 (73)
Multiracial	2 (13)
Hospital role, n (%)	
Registered nurse	6 (40)
Critical care fellow	1 (7)
Hospitalist	4 (27)
Nurse practitioner	2 (13)
Physician assistant	2 (13)
Years in current role, median (IQR)	3.5 (1.5-5)
Years of inpatient experience, median (IQR)	4 (3.5-7)

Table 2. Themes derived from interviews with clinicians about their preferences for a sepsis-focused predictive clinical decision support system with definitions and examples.

Theme and definition	Examples
Clinician perceptions	
Positive sentiments	
Statements made that reflect positive feelings or opinions about predictive information. Includes statements that describe building or already having trust in predictive information.	• Helpful when the data are not giving a clear picture or unsure of course of action • When a prediction is tied to a specific intervention or relevant clinical decision-making • Clinician education efforts to explain relevant studies, model validation, and predicted outcomes
Negative sentiments	
Statements made that reflect negative feelings or opinions about predictive information. Includes statements that describe losing trust or having mistrust in predictive information.	• A clinician feeling like they want to go based off other their own gestalt rather than trusting an alert without a clear explanation • Frequently dismissing false positive alerts
Previous experiences	
Previously deployed sepsis alerts	
Discussion of EHR[a]-based sepsis-specific alerts that were previously or are implemented in the health system.	• Two prior iterations of a sepsis-specific EWS[b] (EWS 1.0 and 2.0)
Risk scores and predictions	
Discussion of bedside clinical risk scores that clinicians have experience using. This includes predictive information for both sepsis and other clinical conditions.	• Wells' criteria • SIRS[c] • Quick sequential organ failure assessment • Ranson score • CHA_2DS_2-VASc
Desired characteristics	
Supporting information	
Clinical information contained in a potential alert to illustrate the reasons an outcome may occur. Additionally, any resources that would be available or linked within an alert	• Vital sign trends • Quantitative presentation of risk information • Links to antibiotic decision tree or antibiotic stewardship info to guide treatment decisions
Platform delivery	
The interface, vector for delivery, timing of delivery, and placement of a potential alert.	• Text alerts • BPA[d]
Predictions	
Clinical outcomes that may occur in patients who are at risk or have developed sepsis and that would be helpful to predict at the bedside.	• Mortality • Transfer to intensive care unit • Development of sepsis or septic shock
Potential utility	
Audience	
Discussion of the best recipients to target for receiving a potential alert to render it useful rather than being dismissed.	• More useful for novice practitioners • Less useful for nurses who do not put in orders • Clinicians changing services, infrequently rotating on a service
Clinical impact	
The potential impacts of an alert on the course of clinical care.	• Change decisions about if and when to initiate broad-spectrum antibiotics • Clarifying to users how might clinical care change based on an alert?

[a]EHR: electronic health record.

[b]EWS: early-warning system.

[c]SIRS: systemic inflammatory response system.

[d]BPA: best practice alert.

Clinician Perceptions

Clinicians' perceptions of predictive information, including positive and negative sentiments, are closely linked to their trust in an alert. Many of the study participants shared that a clinician's knowledge of the background and development of a CDS system contributes to trust in that system when deployed in a clinical setting:

> *I'm someone that attends grand rounds and evidence-based medicine presentations, so I would be a participant in something like that. And so, that would be a useful way to get the information out. Any information that helps to determine how it was made, I think, whether it's from studies done at the hospital, or from evidence taken—reviewed from different articles. I think things like that really do carry a lot of weight, especially if there's something in an algorithm that doesn't immediately intuitively make—isn't what you thought it would be. It's helpful to have information to understand how you got to that point, because then you learn something.* [Advanced practice provider 3]

Participants described additional methods for clinician engagement and education that would foster acceptance of a new alert, such as presenting information about the alert's background during pre-existing information sharing venues, distributing previous research, and clinical leadership providing educational opportunities for those using the alert on the floor:

> *Anytime something new is started on the hospital we always have huddles...sometimes teams will come and say we have this new product that we're implementing or there's a new protocol for sepsis, so teams will kind of round...And then also just working with the leadership team because they can—they really disseminate, every week a lot of the leadership teams will disseminate like new things. So the CNS [Clinical Nurse Specialist] and CPL [Clinical Practice Leader] teams do a great job of kind of educating the nurses. So I feel like if there was to be something newly implemented, those teams specifically will help you make a plan on what is possible.* [Bedside nurse 3]

Clinicians' assessments may not align with the predictions made on the alert, thereby diminishing trust in the algorithm. Clinicians viewed alerts negatively and felt that they may be inaccurate or fire too frequently:

> *So, I would like to see something that doesn't trigger every single time there's a small heart rate change because maybe my patient just went for a walk with physical therapy or is getting out of bed and their heart rate is 120, but they're also getting out of bed for the first time in two weeks.* [Bedside nurse 4]

In addition, alerts were viewed as unhelpful when they were not actionable, with a clear next step toward patient care:

> *Because, honestly, sometimes there are things in EPIC right now, obviously, that pop up. And they ask you if you would consider sepsis in this patient. But to my knowledge, right now it just shows you a bunch of vital signs and, if you've been following the patient for a couple days and you're pretty confident in your treatment plan, you just kind of hit, okay, no thanks, or no suspected sepsis, and just kind of move on. I mean, I can't tell you how many times I've probably just went ahead and hit that button just to get it off the screen because I'm trying to do something else.* [Advanced practice provider 1]

Previous Experiences

Previous experiences with predictive information in the form of clinical risk scores and alerts embedded in the EHR were common among the participants. Clinicians had varied experiences with such risk predictions, which included sepsis-specific tools such as systemic inflammatory response syndrome, quick sequential organ failure assessment, and EHR-embedded predictive alerts. Some clinicians have highlighted risk scores based on their usefulness or lack thereof in making clinical decisions:

> *I do think it comes down to the whole question of, like, modified SOFA versus qSOFA criteria or even SIRS in the sense that if the patient shows more than like three or four or even five things it's not going to be very valuable to me if I'm the one who's responsible for recognizing it in the absence of an alert coming up. And that's why no one uses the SOFA criteria is because there's like nine different things and I can't remember them.* [Physician 4]

Several clinicians had personal interactions with previously deployed early-warning systems and drew on these experiences to reflect on the usefulness of sepsis-specific CDS systems:

> *I don't know if this is still available and in EPIC, and just not at Penn anymore, but I know there used to be a sepsis—a screener tool based on the data that used to pop up. I don't know if you're familiar with that. I remember entering certain vital signs and getting a notification that this patient is at risk for sepsis. But I think that was helpful in identifying trends early and that are so slight that nursing probably wouldn't think anything of. I think they did away with it, at one point, just because of how frequently it was going off and it wasn't always 100 percent accurate.* [Bedside nurse 2]

Overall, these alerts were described as unfavorable because of the perceived high frequency and low accuracy, in addition to disrupting the usual clinical workflow.

Desired Characteristics

Clinicians described the desired characteristics of a potential predictive model and alert, including its predictions of clinical

outcomes, supporting information regarding the patient's status, and the platform on which it is delivered:

> *Like immunosuppressed status, if they were on immunosuppressive medications, if they have underlying malignancy, if they have risk factors for infection, age, if they are community dwellers or if they're coming from nursing homes or care facilities, if they're hospitalized–already hospitalized patients. Those are some of the kinds of things I would be thinking about.* [Advanced practice provider 2]

Participants sought an alert that would deliver a prediction about a specific clinical outcome in the near future. Risk of mortality, future requirement for intensive care unit transfer, need for antibiotics or mechanical ventilation, and being at risk for sepsis or septic shock were suggested as helpful clinical outcomes to present in an alert. Clinicians also expressed a preference for numeric data to contextualize a patient's risk for a specific outcome. Although there was no agreement regarding the specific thresholds, many clinicians felt that plain language such as "your patient has an 80% chance of developing septic shock in the next 24 hours" would be clinically actionable:

> *If someone told me this patient, based—has a likelihood of, I don't know, greater than 30 percent in-hospital mortality I might be more likely to pull the trigger on antibiotics....Other criteria – I guess you could say chance of discharge to home versus a rehab facility. In my mind, that shows whether we caught the infection quick enough so that their level of debilitation was less and they can go home or if they were so debilitated because we waited so long that they now need to go to a physical therapy skilled nursing facility for a couple of weeks. Other things, I think percent chance that they have to go to the ICU, for example, because maybe right now when I recognize the infection and sepsis you don't need to go to the ICU, but if someone told me this patient has a 33 percent chance of going to the ICU, I might be inclined to act quicker.* [Physician 3]

In addition, it is important for clinicians to present appropriate supporting information regarding a patient's clinical presentation to contextualize risk prediction. Trends such as for fever, heart rate, and laboratory values were of particular importance to clinicians who prioritized tracking a patient's trends over time rather than viewing an isolated value. Additional desired alert features included trends in laboratory results and vital signs, reduction of false alarms, and explanatory content about what variables drive a model's prediction:

> *Like again, to incorporate the idea of a trend. Like how new is this abnormality, and abnormality meaning combining not just vital signs, but lab values and orientation status and all of these things that come with sepsis.* [Bedside nurse 4]

> *I guess, to me, the trends are so helpful that I know where to find them. But it's not super intuitive, in Epic when you see a white count, you kind of need to scroll to see what the white count has been.* [Advanced practice provider 4]

There was a consensus among participants that unobtrusive alert delivery methods would be the best. Clinicians favored an easy-access location such as the patient summary tab or chart advisories section in the EHR or a flag highlighting the value of concern when viewing laboratory results. Alerts that presented as a hard stop (such as a best practice advisory) by requiring acknowledgment before navigating other areas of the electronic medical record were viewed unfavorably. Alerts with *soft* stops—notifications not requiring user acknowledgment—that provide guidance for future action and could be accessed at the clinicians' convenience, were preferred:

> *Just stick it in with the chart advisories. It pops up every once in a while. Not every single time you open EPIC, but the first time after it generates the chart advisory, you have to acknowledge it, give a reason like provider notified. And then it goes away for a while. It doesn't keep coming up every time you open the goddamn chart.* [Bedside nurse 5]

Potential Utility

Clinicians felt strongly that an alert's utility would be integral to its success. We identified the delivery of relevant information to the appropriate audience and positive impacts on clinician care as 2 important factors contributing to the perceived utility of a CDS system. Participants from all clinical roles suggested directing alerts to specific units where sepsis is not as common or to clinicians who do not rotate frequently in wards where sepsis is seen:

> *For those of us who see–as internal medicine people who treat infections all the time, I don't think it's like that helpful unless you're pretty novice. Where I think this is most helpful is to the person who does not usually take care of these types of patients, so like to some degree surgery or folks who are in ultra-subspecialties that would always defer this to someone.* [Advanced practice provider 2]

Some nurses felt that an alert would be more beneficial for the clinician responsible for placing orders than for the entire team, as the clinician is less likely to observe incremental temporal changes in the patient yet is charged with making antibiotic administration decisions:

> *So, what I'm saying is it would be more helpful for providers who are not with the patient at bedside, and who sometimes never actually even see the patient with their own eyes. So, especially on night shift. On night shift, the provider does not come see the patients unless there's some clinical indication that the nurse has brought to their attention.* [Bedside nurse 5]

There was agreement among clinicians regarding the need for an actionable alert that directly affects patient care, such as more frequent monitoring, ordering additional or repeat testing, and initiation of a sepsis protocol or antibiotics:

> *A big thing is, of course, appropriate use of antibiotics to make sure that we're not over treating patients. And so, maybe something that reminds you to reevaluate your antibiotic use at the 24-hour mark is something that could be helpful. Because a lot of times*

> *when we're not sure what's going on, we do add a lot on initially, and then we get a lot more information and then we –it is appropriate to start peeling things back. Other times it's not, and someone –and we don't find out what's going on, and so we continue to treat someone empirically. But something like that could be helpful to prompt you to really think about the antibiotic decisions that you're making and to think about antimicrobial stewardship.* [Advanced practice provider 3]

In addition, physicians and APPs recognized the potential of alerts to have positive effects on antibiotic stewardship, while nurses noted being able to use a previous CDS tool to facilitate advocating for patients and prompting conversations with the larger care team:

> *When this alert would go off, yes, you had to notify the charge nurse and the team member had to come to the bedside right away and you had to like make a plan and say like, okay, we're going to draw cultures. Now I feel like – we don't have an alert, but the nurses and the whole team does a really good job of like alerting the team, making sure it's like a phone call and advocating for blood cultures and all of those things now...So I feel like we've had a really good – since they stopped that alert, I haven't seen – I'm unaware of like the nursing not notifying the team and advocating for the right things or the team not starting things appropriately.* [Bedside nurse 3]

Discussion

Principal Findings

In preparation for building and implementing a predictive sepsis CDS alert in an academic health system, we interviewed physicians, APPs, and nurses about their experiences and perceptions of predictive CDS systems in clinical workflows around patients suspected of sepsis. We identified themes in these interviews that offer insights into strategies to increase the likelihood of adoption, increase clinical effectiveness, and establish trust in an alert among hospital-based clinicians. These findings have several implications for developing sepsis-focused decision support tools and providing guidance for the creation of trusted and actionable CDS systems.

Clinician Perceptions

Participants expressed interest in clinician engagement and educational activities regarding a predictive CDS system before deployment. Specifically, participants expressed an interest in education on how the model was developed, what specific factors went into the predictions, and how to interact with a predictive alert. These findings complement previous work, suggesting that training and education on the growing presence of artificial intelligence in health care could extend to the organizational level to increase machine learning literacy in clinical staff and overcome some of the barriers in CDS adoption [9,18]. Interviewed clinicians who were either previously exposed to or educated on CDS model development had a much more favorable perception of the system. Opportunities for education and interaction between clinicians and the CDS development team should be a part of any new CDS system integration Previous evidence shows a lack of coordinated implementation strategies lowers the likelihood of adoption of sepsis predictive alerts in multiple previous CDS explorations [19].

These findings highlight the benefits of a prospective assessment strategy, rather than a retrospective one, because the impact of a CDS is greatly influenced by stakeholder adoption and buy-in [20,21]. This is in contrast to recent qualitative work in this field that has focused primarily on clinician perceptions of existing and previously implemented sepsis alerts [8,9]. Although still useful, these retrospective analyses are limited to the characteristics of alerts that were already developed and implemented.

Previous Experiences

Clinicians' frustrations were evident in "hard-stop" alerts that require user action or acknowledgment, commonly citing experiences with best practice advisories that were viewed as unhelpful and contributing to workflow delays. The effects of such undesirable CDS characteristics can lead to alert fatigue and interruptions in cognitive and clinical workflows, which in turn leads to delays in the initiation of antibiotics [22-24]. A total of 2 previous mixed methods studies evaluating clinician perceptions of a previously deployed sepsis early warning system in our health system reported low clinical relevance and low likelihood of affecting clinical patient management [8,17]. These alerts, especially those contradicting a clinician's impression of a patient, were met with a lukewarm response from clinicians, as teams were required to meet and discuss the patient within a short period, similarly interfering with the normal clinical workflow on the ward [17].

Desired Characteristics

Importantly, we identified preferences for alert characteristics that are unique to both bedside nurses and clinicians responsible for ordering antibiotics. Participants in the nursing group spoke about how sepsis-specific predictive CDS systems might support the need both to advocate for patients and to relay critical information to other clinicians who make decisions on ordering antibiotics and other diagnostic tests. Preferences expressed by bedside nurses extend and complement the findings of previous studies that showed nursing preferences for alerts that provide timely care recommendations, highlight treatment protocols, and address a patient's condition, rather than those that emphasize regulatory guidelines [25]. Although bedside nurses are not consistently included as recipients of sepsis-based electronic alerts [26,27], these findings underscore their importance in caring for patients suspected of sepsis and how a sepsis CDS system might address some of the challenges that they face in relation to information gathering and team communication.

Physicians and APPs identified specific elements of a patient's EHR data and history that could help them in their treatment decisions. A patient's past antibiotic history, microbiological data, and comorbidities, such as underlying malignancies and immunosuppressed status, were all recognized by multiple clinicians as especially relevant in determining appropriate

antibiotics to be used during treatment initiation. Clinicians also expressed a preference for specific antibiotic guidance to balance therapeutic efficacy and stewardship. A sepsis-focused CDS should provide easy access to these EHR data elements for clinicians to facilitate decision-making based on alerts. Notably, these treatment-focused data elements were not highlighted in the nurses' responses, which were, in contrast, more focused on immediate patient care concerns and communicating risks to the clinicians responsible for placing orders. Although some of our findings, such as the desire to see data trends in the CDS system, are consistent with previously reported results [17], our results extend prior work in this area by identifying additional features of both the alert itself and the health system's approach to engaging clinicians before deploying an alert that can inform the planning of future development and deployment of sepsis-specific CDS systems.

Potential Utility

Our participants desired an easy-to-digest alert that was, most importantly, accessible, unobtrusive, and believed to be clinically accurate. The acceptability of sepsis CDS relies on both its prediction accuracy and its presentation of information in a readily interpretable manner [28]. There is a growing body of evidence investigating the importance of human factors in CDS design, ranging from alert type to textual and graphical displays of information [25,28]. Clinicians made numerous suggestions along these lines, including displaying a "flag" or marker in the "summary" tab of a patient's chart to assist the clinician in recognizing an issue without interrupting usual workflows. Nurses specifically highlighted the advantage of even a small visual signal to review the patient's trending information, allowing them to take a closer look at patient temporal data in cases that they otherwise may not have.

A sepsis CDS system represents a complex interaction between technological factors such as flagged alerts and display of information, and social factors such as communication between nurses and physicians. In designing and implementing such a system, detailed aspects of both technology and its use in the hospital setting should be considered. Sociotechnical theory takes a measured approach to these interactions and, over the years, multiple frameworks have been developed to provide a conceptual structure in system design. The Systems Engineering Initiative for Patient Safety model [29] emphasizes the interaction between people, technology, environment, and organization [30]. Another example, Sitting and Singh's health care sociotechnical framework [31,32], contains 8 dimensions detailing computing infrastructure, the human-computer interface, clinical content, and organizational policies that are used to assess barriers and facilitators when implementing systems. The health care sociotechnical framework is particularly applicable in sepsis CDS design. Granular technology details can be described in the framework, such as the distinction between a *soft stop* and *hard-stop* alert as discussed by our clinicians and the specific steps one would take to acknowledge such an alert. Considering these distinctions early in system design will increase the end user acceptance and utility of a sepsis CDS. The social dimensions of the framework describe the interactions between the clinical staff *users* and those who design, develop, and implement these systems. Explorations such as the one we conducted are the first step in the successful development of a sepsis CDS that considers user perceptions, past experiences, and desired alert characteristics with a high likelihood of clinical utility.

Limitations

Our study had several limitations. First, all participants had <10 years of experience in an inpatient setting. Thus, this study does not reflect the preferences of more experienced clinicians with distinct practice patterns or different experiences with predictive CDS systems. Second, we only interviewed clinicians from a single health system, and the findings may not be generalized to other health systems with different patient populations, cultures, EHR systems, and previous experiences with sepsis-focused predictive CDS systems. However, illustration rather than generalizability is the intended goal of qualitative research, and the approach outlined here provides a framework for eliciting clinician preferences locally and prospectively, which can be adapted elsewhere.

Conclusions

This study provides a more detailed understanding of clinician preferences for predictive alerts to assist in the care of patients with a potential sepsis diagnosis. Physicians, APPs, and bedside nurses desire a sepsis-focused predictive sepsis CDS system that is trusted, unobtrusive, and viewed as actionable at the bedside. Opportunities exist for sepsis CDS systems not only to improve diagnosis and treatment decisions but also to facilitate communication in a multidisciplinary team setting. Eliciting stakeholder feedback and identifying preferences for predictive alerts before model development offers an opportunity to engage in clinician education and outreach, which may improve the acceptability and adoption of future sepsis CDS systems.

Acknowledgments

The authors would like to thank Jon Pomeroy for assisting in creating simulated electronic health record chart examples in EPIC (Epic Systems Corporation).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Nursing and physician interview guides used during semistructured interviews.

[DOCX File , 26 KB - humanfactors_v9i4e36976_app1.docx]

References

1. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 2001 Jul;29(7):1303-1310. [doi: 10.1097/00003246-200107000-00002] [Medline: 11445675]
2. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 2003 Apr 17;348(16):1546-1554. [doi: 10.1056/NEJMoa022139] [Medline: 12700374]
3. Amland RC, Hahn-Cover KE. Clinical decision support for early recognition of sepsis. Am J Med Qual 2019;34(5):494-501 [FREE Full text] [doi: 10.1177/1062860619873225] [Medline: 31479290]
4. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 2006 Jun;34(6):1589-1596. [doi: 10.1097/01.CCM.0000217961.75225.E9] [Medline: 16625125]
5. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 2016 Sep 30;4(3):e28 [FREE Full text] [doi: 10.2196/medinform.5909] [Medline: 27694098]
6. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019 Mar;170:1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]
7. Fagerström J, Bång M, Wilhelms D, Chew MS. LiSep LSTM: a machine learning algorithm for early detection of septic shock. Sci Rep 2019 Oct 22;9(1):15132 [FREE Full text] [doi: 10.1038/s41598-019-51219-4] [Medline: 31641162]
8. Ginestra JC, Giannini HM, Schweickert WD, Meadows L, Lynch MJ, Pavan K, et al. Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. Crit Care Med 2019 Nov;47(11):1477-1484 [FREE Full text] [doi: 10.1097/CCM.0000000000003803] [Medline: 31135500]
9. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a machine learning system into clinical workflows: qualitative study. J Med Internet Res 2020 Nov 19;22(11):e22421 [FREE Full text] [doi: 10.2196/22421] [Medline: 33211015]
10. Pertiwi AA, Fraczkowski D, Stogis SL, Lopez KD. Using heuristic evaluation to improve sepsis alert usability. Crit Care Nurs Clin North Am 2018 Jun;30(2):297-309. [doi: 10.1016/j.cnc.2018.02.011] [Medline: 29724447]
11. Miller K, Kowalski R, Capan M, Wu P, Mosby D, Arnold R. Assessment of nursing response to a real-time alerting tool for sepsis: a provider survey. Am J Hosp Med 2017;1(3):2017.021 [FREE Full text] [doi: 10.24150/ajhm/2017.021] [Medline: 30854401]
12. Sidebottom AC, Collins B, Winden TJ, Knutson A, Britt HR. Reactions of nurses to the use of electronic health record alert features in an inpatient setting. Comput Inform Nurs 2012 Apr;30(4):218-228. [doi: 10.1097/NCN.0b013e3182343e8f] [Medline: 22045117]
13. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
14. Prescott HC, Iwashyna TJ. Improving sepsis treatment by embracing diagnostic uncertainty. Ann Am Thorac Soc 2019 Apr;16(4):426-429 [FREE Full text] [doi: 10.1513/AnnalsATS.201809-646PS] [Medline: 30883190]
15. Timmermans S, Tavory I. Theory construction in qualitative research. Sociol Theory 2012 Sep 10;30(3):167-186. [doi: 10.1177/0735275112457914]
16. Halpin M, Richard N. An invitation to analytic abduction. Methods Psychol 2021 Dec;5:100052 [FREE Full text] [doi: 10.1016/j.metip.2021.100052]
17. Guidi JL, Clark K, Upton MT, Faust H, Umscheid CA, Lane-Fall MB, et al. Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center. Ann Am Thorac Soc 2015 Oct;12(10):1514-1519. [doi: 10.1513/AnnalsATS.201503-129OC] [Medline: 26288388]
18. Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. NPJ Digit Med 2019 Apr 26;2:28 [FREE Full text] [doi: 10.1038/s41746-019-0106-0] [Medline: 31304375]
19. Joshi M, Ashrafian H, Arora S, Khan S, Cooke G, Darzi A. Digital alerting and outcomes in patients with sepsis: systematic review and meta-analysis. J Med Internet Res 2019 Dec 20;21(12):e15166 [FREE Full text] [doi: 10.2196/15166] [Medline: 31859672]
20. Beeler PE, Bates DW, Hug BL. Clinical decision support systems. Swiss Med Wkly 2014 Dec 23;144:w14073 [FREE Full text] [doi: 10.4414/smw.2014.14073] [Medline: 25668157]
21. Downing NS, Wang C, Gupta A, Wang Y, Nuti SV, Ross JS, et al. Association of racial and socioeconomic disparities with outcomes among patients hospitalized with acute myocardial infarction, heart failure, and pneumonia: an analysis of within- and between-hospital variation. JAMA Netw Open 2018 Sep 07;1(5):e182044 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2044] [Medline: 30646146]

22. Sakata KK, Stephenson LS, Mulanax A, Bierman J, Mcgrath K, Scholl G, et al. Professional and interprofessional differences in electronic health records use and recognition of safety issues in critically ill patients. J Interprof Care 2016 Sep;30(5):636-642 [FREE Full text] [doi: 10.1080/13561820.2016.1193479] [Medline: 27341177]
23. The Lancet Respiratory Medicine. Crying wolf: the growing fatigue around sepsis alerts. Lancet Respir Med 2018 Mar;6(3):161. [doi: 10.1016/S2213-2600(18)30072-9] [Medline: 29508700]
24. Kohn R, Ginestra J, Hubbard R, Auriemma C, Patel M, Anesi G, et al. 1406: Association of time of day with antimicrobial start for ward patients with hospital-acquired sepsis. Crit Care Med 2022 Jan;50(1):705. [doi: 10.1097/01.ccm.0000811948.98477.e7]
25. Long D, Capan M, Mascioli S, Weldon D, Arnold R, Miller K. Evaluation of user-interface alert displays for clinical decision support systems for sepsis. Crit Care Nurse 2018 Aug;38(4):46-54 [FREE Full text] [doi: 10.4037/ccn2018352] [Medline: 30068720]
26. Hooper MH, Weavind L, Wheeler AP, Martin JB, Gowda SS, Semler MW, et al. Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit*. Crit Care Med 2012 Jul;40(7):2096-2101 [FREE Full text] [doi: 10.1097/CCM.0b013e318250a887] [Medline: 22584763]
27. Semler MW, Weavind L, Hooper MH, Rice TW, Gowda SS, Nadas A, et al. An electronic tool for the evaluation and treatment of sepsis in the ICU: a randomized controlled trial. Crit Care Med 2015 Aug;43(8):1595-1602 [FREE Full text] [doi: 10.1097/CCM.0000000000001020] [Medline: 25867906]
28. Schubel L, Mosby DL, Blumenthal J, Capan M, Arnold R, Kowalski R, et al. Informatics and interaction: applying human factors principles to optimize the design of clinical decision support for sepsis. Health Informatics J 2020 Mar;26(1):642-651 [FREE Full text] [doi: 10.1177/1460458219839623] [Medline: 31081460]
29. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1:i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
30. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033 [FREE Full text] [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
31. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
32. Or C, Dohan M, Tan J. Understanding critical barriers to implementing a clinical information system in a nursing home through the lens of a socio-technical perspective. J Med Syst 2014 Sep;38(9):99. [doi: 10.1007/s10916-014-0099-9] [Medline: 25047519]

Abbreviations

APP: advanced practice provider
CDS: clinical decision support
EHR: electronic health record

Edited by A Kushniruk; submitted 01.02.22; peer-reviewed by C Or, AF Näher, N Yung; comments to author 14.05.22; revised version received 05.08.22; accepted 06.08.22; published 21.10.22.

Please cite as:
Silvestri JA, Kmiec TE, Bishop NS, Regli SH, Weissman GE
Desired Characteristics of a Clinical Decision Support System for Early Sepsis Recognition: Interview Study Among Hospital-Based Clinicians
JMIR Hum Factors 2022;9(4):e36976
URL: https://humanfactors.jmir.org/2022/4/e36976
doi:10.2196/36976
PMID:36269653

Original Paper

A Mobile App Leveraging Citizenship Engagement to Perform Anonymized Longitudinal Studies in the Context of COVID-19 Adverse Drug Reaction Monitoring: Development and Usability Study

Marzia Di Filippo[1], PhD; Alessandro Avellone[1], PhD; Michael Belingheri[2], PhD; Maria Emilia Paladino[2], PhD; Michele Augusto Riva[2], PhD; Antonella Zambon[1], PhD; Dario Pescini[1], PhD

[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[2]School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

Corresponding Author:
Dario Pescini, PhD
Department of Statistics and Quantitative Methods
University of Milano-Bicocca
Via Bicocca degli Arcimboldi 8
Milan, 20126
Italy
Phone: 39 0264485835
Email: dario.pescini@unimib.it

Abstract

Background: Over the past few years, studies have increasingly focused on the development of mobile apps as complementary tools to existing traditional pharmacovigilance surveillance systems for improving and facilitating adverse drug reaction (ADR) reporting.

Objective: In this research, we evaluated the potentiality of a new mobile app (vaxEffect@UniMiB) to perform longitudinal studies, while preserving the anonymity of the respondents. We applied the app to monitor the ADRs during the COVID-19 vaccination campaign in a sample of the Italian population.

Methods: We administered vaxEffect@UniMiB to a convenience sample of academic subjects vaccinated at the Milano-Bicocca University hub for COVID-19 during the Italian national vaccination campaign. vaxEffect@UniMiB was developed for both Android and iOS devices. The mobile app asks users to send their medical history and, upon every vaccine administration, their vaccination data and the ADRs that occurred within 7 days postvaccination, making it possible to follow the ADR dynamics for each respondent. The app sends data over the web to an application server. The server, along with receiving all user data, saves the data in a SQL database server and reminds patients to submit vaccine and ADR data by push notifications sent to the mobile app through Firebase Cloud Messaging (FCM). On initial startup of the app, a unique user identifier (UUID) was generated for each respondent, so its anonymity was completely ensured, while enabling longitudinal studies.

Results: A total of 3712 people were vaccinated during the first vaccination wave. A total of 2733 (73.6%) respondents between the ages of 19 and 80 years, coming from the University of Milano-Bicocca (UniMiB) and the Politecnico of Milan (PoliMi), participated in the survey. Overall, we collected information about vaccination and ADRs to the first vaccine dose for 2226 subjects (60.0% of the first dose vaccinated), to the second dose for 1610 subjects (43.4% of the second dose vaccinated), and, in a nonsponsored fashion, to the third dose for 169 individuals (4.6%).

Conclusions: vaxEffect@UniMiB was revealed to be the first attempt in performing longitudinal studies to monitor the same subject over time in terms of the reported ADRs after each vaccine administration, while guaranteeing complete anonymity of the subject. A series of aspects contributed to the positive involvement from people in using this app to report their ADRs to vaccination: ease of use, availability from multiple platforms, anonymity of all survey participants and protection of the submitted data, and the health care workers' support.

(JMIR Hum Factors 2022;9(4):e38701) doi:10.2196/38701

KEYWORDS

ADR reporting; adverse drug reaction–reporting systems; mobile apps; longitudinal studies; COVID-19 vaccination campaign; COVID-19; vaccine; apps; adverse drug reaction; pharmacovigilance; anonymity

Introduction

Pharmacovigilance is defined by the World Health Organization (WHO) as "the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem" [1]. Monitoring the safety of medicines is fundamental because many previously undetected adverse drug reactions (ADRs) during clinical trials may occur when patients are exposed to drugs. ADRs refer to any undesired effect emerged during treatment with a given pharmaceutical product [2].

We performed a literature analysis to review the currently adopted methods for ADR reporting. In particular, we searched PubMed for papers published in the past 10 years using the following query: "adverse drug reactions" OR "adverse drug reaction reporting" OR "adverse drug reaction reporting systems" (query performed on June 22, 2022). The considered papers included clinical trials, meta-analyses, randomized controlled trials, and reviews. This search identified 131 studies. By excluding papers that, after a full text or abstract screening, seemed too descriptive and generic or focused on just presenting the incidence or prevalence of ADRs, the literature survey performed was limited to 28 eligible papers. We found 9 other papers that met the considered criteria. We grouped the studies based on the system used to conduct ADR surveillance.

Of the total 39 papers found, 7 (17.9%) [3-9] are based on traditional safety surveillance. This approach relied on data collected through passive reports of ADRs from consumers and health providers, medical literature, and observational databases [10]. This way, any active measure is taken to search for ADRs, except for voluntary and spontaneous reports by health care providers or vaccinees [11,12]. Although passive reporting is still used, thanks to the possibility of identifying rare and unexpected ADRs, its ease of implementation, and relatively low cost, it can be limited by underreporting, lack of information, and difficulty in determining ADR rates [12,13]. Moreover, traditional surveillance of ADRs after drug exposure is slow and patchy. Indeed, once reported by patients or health care professionals, ADRs are first assessed by drug experts and pharmaceutical companies, and then the result of the assessment is passed on to government agencies. This path may entail underreporting issues, with considerable data loss.

Substantial delays may occur between occurrence and discovery of ADRs, because reports received by government agencies are often released with a delay of months or even years due to the period required for proper assessment of ADR data. Further limitations of traditional ADR surveillance relate to the fact that ADR reporting is voluntary, and several studies have shown that as many as 90% of serious ADRs end up being unreported [2,14]. Moreover, several studies have reported low awareness of the national public ADR reporting systems within the general population, with a portion of the public not even being aware of them. This adds up to a problem of declining reporting by physicians that has been linked to the belief that it is only necessary to report serious or unexpected ADRs or that a single observed ADR could not contribute to medical knowledge or that only ADRs derived from a certain causal relationship with the use of a particular drug should be reported [10,15].

In response to these limitations, ADR surveillance is also conducted by applying a participant-centered active approach that, contrary to passive reporting, involves an intentional searching for ADRs via continuous and organized contact with health care providers or other relevant reporting sources, such as hospitals, laboratories, and patients [11,12].

At the same time, active surveillance requires more resources for a positive outcome. Of the 39 papers, in 16 (41%), active systems falling under this category used digital strategies to facilitate and encourage ADR surveillance, which means the adoption of electronic technology for capturing and processing data.

Common participant-centered active surveillance methods are conducted using health diary reporting cards [11], SMS text messages (eg, SmartVax [16], VaxTracker [17], stimulated telephone-assisted rapid safety surveillance [STARSS] [18], FASTMum [19], FAST-Health [20]), and email (eg, VaxTracker [17], Lareb Intensive Monitoring [21], and integrated vaccine surveillance system [IVSS] [22]) to interact with individuals and prompt them to report their ADRs by sending a web link to complete an online questionnaire, interviews collected over the phone (eg, FASTMum [19], IVSS [22], and a project developed in Brazil [13,23,24]), during medical visits [25], or mailing questionnaires [26-29].

Thanks to the spread of internet use, our literature survey suggested increasing usage of web-based platforms to promote ADR reporting. A large proportion of the papers found (16/39, 41%) adopted interactive web-based ADR-reporting tools [30-33] also accessible from mobile devices [34] that linked to the receipt of spontaneous ADR reports [30], to the use of electronic health records to directly report ADRs [35-37], and to spontaneous reports sent through multiple forms of notification (eg, via an ADR portal, telephone, and email [38]) and sent on the online Yellow Card scheme [39-42]. The same explored ways are also used individually or in combination to report ADRs through a paper-based form sent by email or fax, an electronic form sent on a website, or telephone interviews [43,44]. This category of web-based platforms to report ADRs also includes the CANVAS program [45], which relies on reports received through a web-based survey, with an additional telephone follow-up when severe ADRs need to be reported.

The results of the European Union's Innovative Medicines Initiative Web–Recognising Adverse Drug Reactions (WEB-RADR) project, reported in a recent paper [46], suggest that the patient's increased interest in improving knowledge about drug safety, combined with recent technological advances in information communication, make it possible to improve

ADR reporting and patient safety communication using mobile apps.

Over the past few years, studies have increasingly focused on the development and usage of mobile apps as additional tools to pharmacovigilance for improving and facilitating reporting of ADRs, as reported by Defer et al [47], Ahn et al [48], Prakash et al [49], Montastruc et al [50], and de Vries et al [51]. Indeed, in contrast to classical surveillance systems, mobile devices offer a platform for improving the accessibility of data and increasing the speed at which they are transmitted between institutions. These tools lead to real-time systems of pharmacovigilance with the potential of enabling a near-instantaneous transmission of patient safety information through user-friendly and interactive graphics within the app [10]. The extensive usage of smartphones and other mobile devices with internet access leads to the expectation of a greater involvement from the public in using mobile health apps to report suspected ADRs, complementing traditional surveillance systems.

In view of these findings, we searched for available mobile apps with the ability to allow ADR reporting for medical devices, drugs, or vaccines, without searching for apps that are specific for a particular case of study.

Among the tools we found, the MedWatcher app [52] is a web-based and mobile app available to the US public, developed in partnership with the Food and Drug Administration (FDA) Center for Devices and Radiologic Health (CDRH) and available on Apple App Store and Google Play Store for iOS and Android devices, respectively. Through MedWatcher, reports are processed in a secure cloud computing environment, manually reviewed, and then transmitted electronically to the Manufacturer and User Facility Device Experience (MAUDE) database. VigiBIP [50], designed by the pharmacovigilance network in France and available for Android and iOS devices, has the notable functionality to allow reporting of not only data but also photographs and images [50,53]. Similarly, the ADR PvPi app [49], which was developed by the National Coordination Centre-Pharmacovigilance Programme of India, facilitates ADR reporting, enabling document and image attachment by health care professionals as well as consumers. However, it was available just on Android devices from the Google Play Store.

Another available tool is presented by Ahn et al [48], where the system design is accurately presented as consisting of a mobile app, a cloud server, and a dashboard. Patients' data on symptoms and vaccination are captured using a mobile device or a web interface and then sent to the cloud server. Medical personnel then use the web dashboard to receive and analyze the submitted data. This mobile app was developed for both Android and iOS devices.

My eReport France [47] is a free mobile app available for both Android and iOS devices. When individuals use this app, they receive a participant card containing quick response (QR) codes to download the app and submit their ADR report.

The WEB-RADR project developed 3 mobile apps suitable for patients, caregivers, and health care professionals not only to report their ADRs but also to receive up-to-date information and news alerts on selected medicines, check the number of received reports for a particular drug, and view previous ADR reports submitted through the app. These 3 apps, which are free to use for everyone on iOS and Android devices, are the free Yellow Card smartphone app [42] developed in 2015, the Halmed app [54] launched in 2016 to submit ADRs to drugs directly to the Agency for Medicinal Products and Medical Devices of Croatia, and the Lareb app [55] developed in 2016 by the Netherlands Pharmacovigilance Centre Lareb but no longer maintained or available for download.

In the context of the WEB-RADR project, the international WEB-RADR app has been launched under the name Med Safety [56]. Similar to the other WEB-RADR apps, Med Safety is a smartphone app freely available on both iOS and Android devices for reporting ADRs to national competent authorities, keeping track of previously reported information and receiving news about drugs of interest. With the ability to function without an internet connection, individuals can partially create reports and save them for completion later or create and save reports without sending them immediately but subsequently once connectivity is re-established.

In the following cases, no specific details about the app design and features, other than the chosen platform and the declared ability to allow ADR reporting, are available: the CANVAS app [45] and the ADR online app [57], which are available only for iOS devices; the ADR Reporter app [58] and the TMDA Adverse Reactions Reporting Tool app [59], which are available only for Android devices; and the Easypharm app [60] and the UAE RADR app [61], which are available for both Android and iOS devices.

Although the development of mobile apps to report ADRs is continuing to progress, providing valuable knowledge about drug safety, one feature that is not present in any of the reviewed apps in our literature analysis is the ability to perform longitudinal analysis in an anonymous way, a key point to foster citizenship engagement.

Although having valuable features, all the currently available mobile apps lack the ability to collect individual subjects' ADR time courses anonymously. Monitoring the same subjects over time, while guaranteeing complete privacy and anonymity of the submitted information, can be a valuable addition to pharmacovigilance because it can be used to gather longitudinal safety data of a drug. Consequently, the temporal investigation of the same patient can help detect, within a specific clinical context, any temporal patterns or ADR combinations.

On these grounds, we developed, for Android and iOS devices, a new, free app called vaxEffect@UniMiB, tested its capability to enable longitudinal studies by analyzing per subject time series, and evaluated its potential in fostering spontaneous citizenship participation in an ADR data collection campaign. In this work, we specifically presented app features by evaluating the reporting rate of ADRs in a sample of Italian academic subjects vaccinated for COVID-19.

Methods

Study Design

During the COVID-19 health emergency, the University of Milano-Bicocca (UniMiB) participated in a vaccination campaign deploying a hub. The hub administered the first dose between March 5 and 29, 2021, and the second dose between May 31 and June 11, 2021. A third dose was administered at the national level, but the authorities decided to not involve the hub. The campaign involved professors, researchers, adjunct professors, postdocs, PhD students, and the technical administrative staff of UniMiB or of the Politecnico of Milan (PoliMi). Due to the deputed authority emergency plan, the UniMiB hub invited all UniMiB employees and only a third of PoliMi ones, the others being distributed to different hubs.

Ethical Considerations

Ethical approval was obtained from the Ethical Committee of UniMiB (protocol no. 0041302/21).

Data Collection

After receiving ethical approval, vaxEffect@UniMiB was made available on March 24, 2021, from Google Play Store and on March 29, 2021, from Apple App Store. It was advertised to subjects invited to the vaccination campaign at the UniMiB hub.

We started to collect reports sent to vaxEffect@UniMiB, starting from January 1, 2021. Given the health emergency, it was not possible to plan the different phases of the study. To highlight the uncertainties of the situation, we report that the UniMiB hub was not aware at the end of the first dose administration whether the second dose would have been administered by the hub itself. When the third dose was administered at the national level, the hub was not involved, and we decided to test the flexibility of the app by keeping it active to allow respondents to use it on a spontaneous basis. The app was deactivated on February 9, 2022.

At each vaccination wave, the physicians involved in the vaccination campaign explained to attendees the importance of reporting ADRs and delivered a brochure explaining the aim of the research, as well as the guarantee of anonymization of the digital tool. Moreover, corresponding to the beginning of the first 2 vaccine doses' administration campaign, an informative mail was sent to UniMiB employees.

System Overview

Our system consists of a mobile app, a web server, and a Firebase Cloud Messaging (FCM) server (Google Inc), as shown in Figure 1. At initial startup, the mobile app inquiries about the user's health condition, focusing only on aspects relevant to the vaccination campaign, and sends the answers to the application server. Upon every vaccine administration, vaccine data and ADRs occurring within 7 days postvaccination are sent by the patient using the mobile app.

The application server is responsible for receiving all the user data, to store them in a SQL database server, and to remind the patients to submit vaccine and ADR data. The reminders are sent to the patients via push notifications sent to the mobile app by the application server through FCM.

The complete data exchange between the mobile app and the application server is described in Figure 2.

Figure 1. Our system. FCM: Firebase Cloud Messaging.

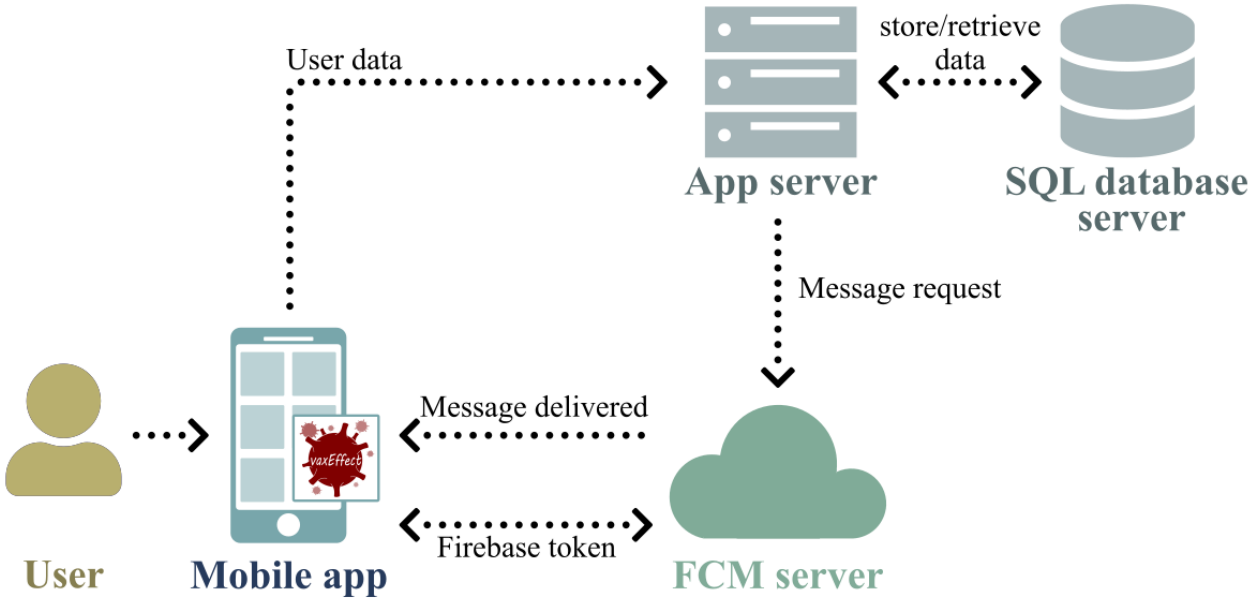
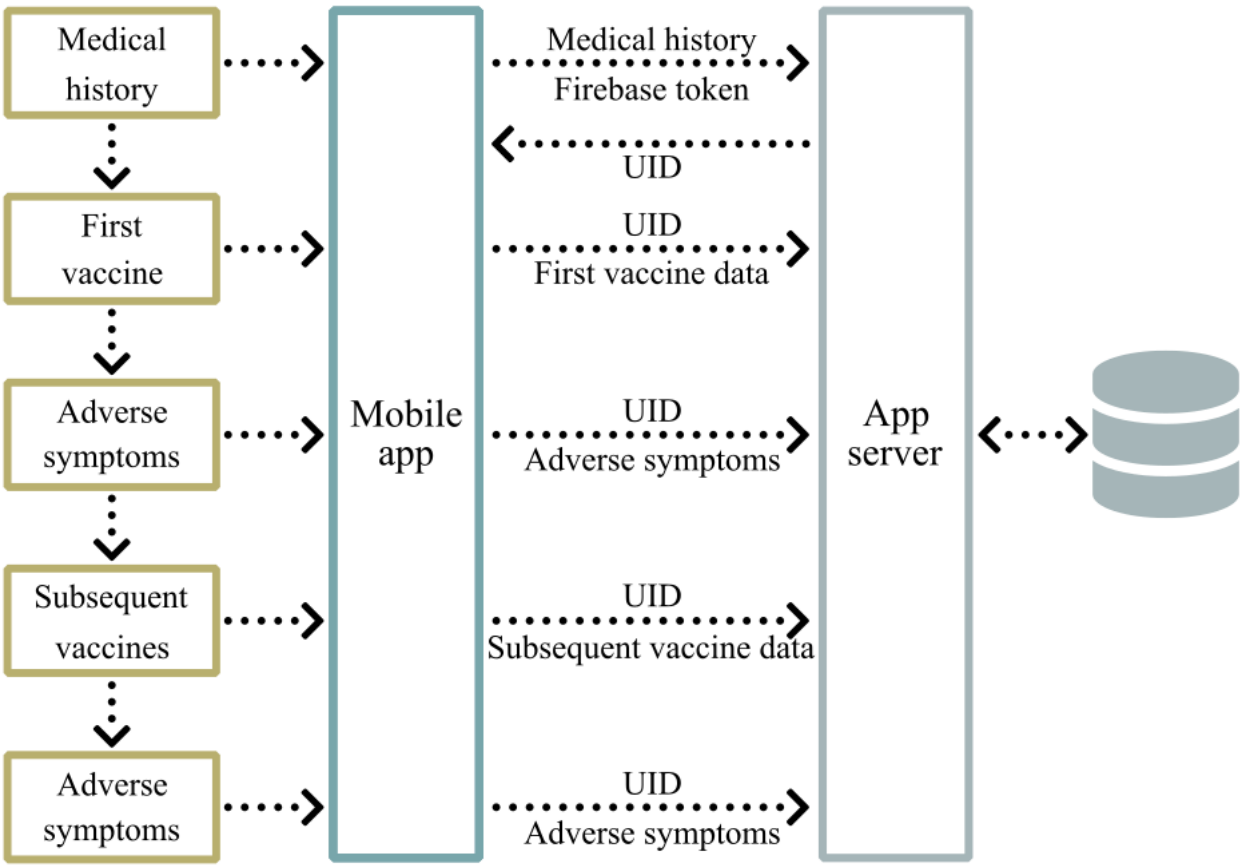


Figure 2. Our system's data exchange. UID: user identifier.



Security by Design

The framework for recording the responses provided by the participants is a one-way client/server architecture. We decided to opt for this type of structure to ensure the safety of the respondents. This way, it was not necessary to store the answers to the questionnaires on the terminal used by the respondents, minimizing the possibility of illegitimate acquisition of information in the event of loss, theft, or fraudulent use of the device.

Furthermore, to ensure the anonymity of the respondents, while preserving the possibility of analyzing the time series relating to the same individual, we decided to generate a unique user identifier (UUID) using an FCM [62] token so that the UUID could not be used to trace the user of the mobile device (the FCM token identifies a single app on a single device and is not associated with any sensitive user data).

Mobile App

The mobile app was developed for both Android and iOS devices using the FCM software development kit (SDK). We specifically used Android Studio (Java language) and XCode (Swift language) for, respectively, the Android and iOS versions. Our app sends data over the web to an application server using an application programming interface (API) gateway. The data are sent using the https protocol and are validated by the application server.

On initial startup of our mobile app, a splash screen is displayed (Figure 3A), and at the same time, the FCM SDK generates a registration token for the client app instance. This token uniquely identifies an app installation and cannot be linked in any way to the device owner, and it is used to receive a push notification from the application server. After the splash screen, a short description of the mobile app (Figure 3B), the privacy agreement (Figure 3C), and a medical history questionnaire (Figure 3D) are displayed. Medical history data and the registration token are sent to the application server. The application server stores the data received in the database and returns a UUID, which is encrypted and saved in the local mobile app file system.

On subsequent startup or after completing the medical history questionnaire, the patient is asked for vaccination data (Figure 4A) and, 7 days later, for ADRs (Figure 4B).

Figure 3. Mobile app first startup. (A) Splash screen. (B) Mobile app description. In this panel, the app is presented, specifying that the user will be asked for vaccination data after the first and, if any, every subsequent vaccination. Seven days after each vaccination, the user will be then asked to submit ADRs. Alerts will be sent to remind the user about the data submission after each moment. In addition, this panel specifies that data are collected on a voluntary basis, ensuring total anonymity of the submitted data. (C) Privacy. In this panel, it is specified that the architecture underlying vaxEffect@UniMiB ensures the privacy of the submitted data, which can be viewed by only the responsible physician and authorized subjects. Moreover, the user needs to confirm to be of legal age, to authorize the processing of the data that will be submitted, and to have read the provided privacy statements. (D) Medical history. At this stage, the user has to complete the medical history questionnaire by inserting demographic data and indicating whether 1 or more chronic diseases among those specified in the "Data collected by the app" section have been ever diagnosed. In this panel, an example of the data that the user needs to submit at this stage is provided by asking whether the user suffers from diabetes, bronchitis, or asthma. ADR: adverse drug reaction.

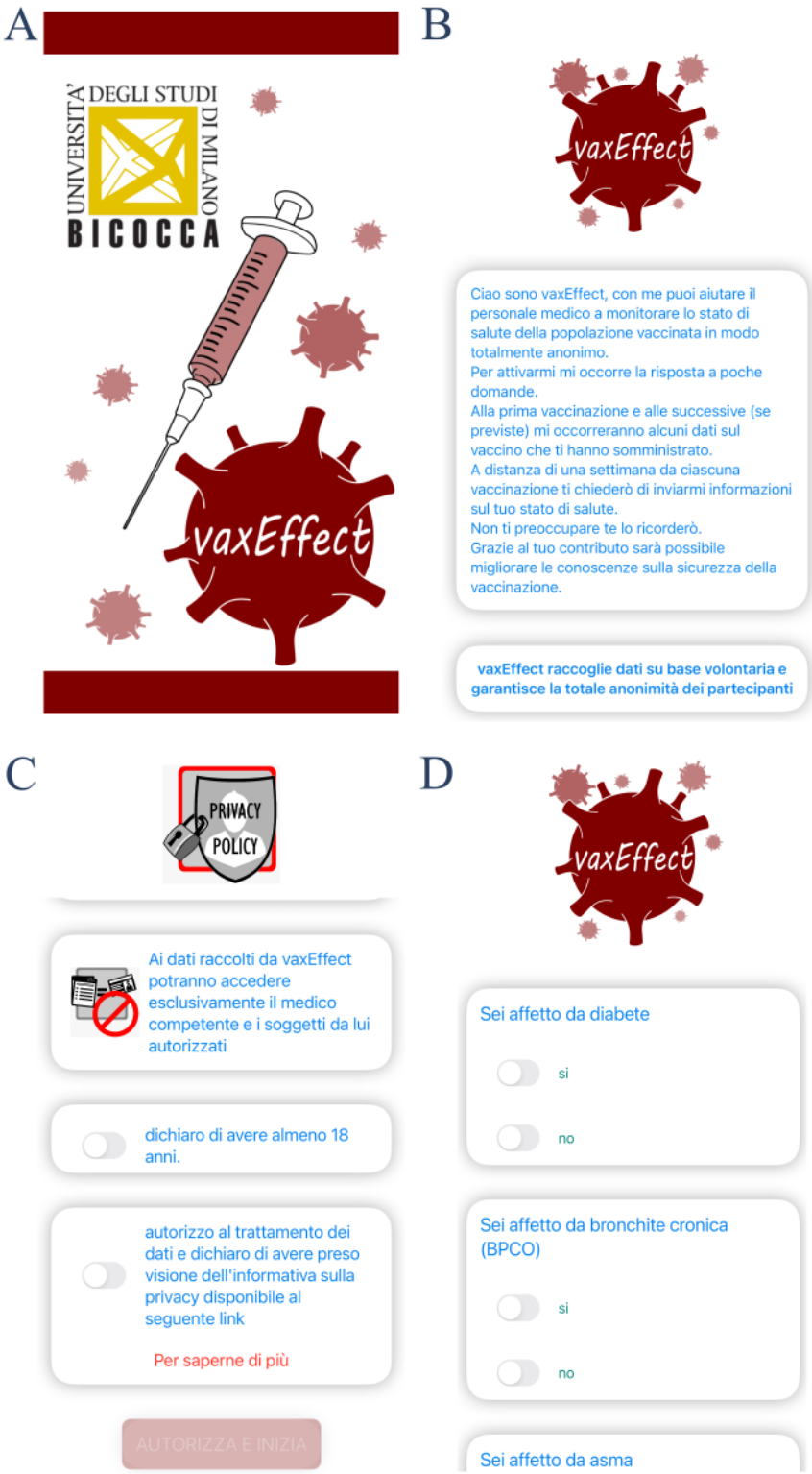
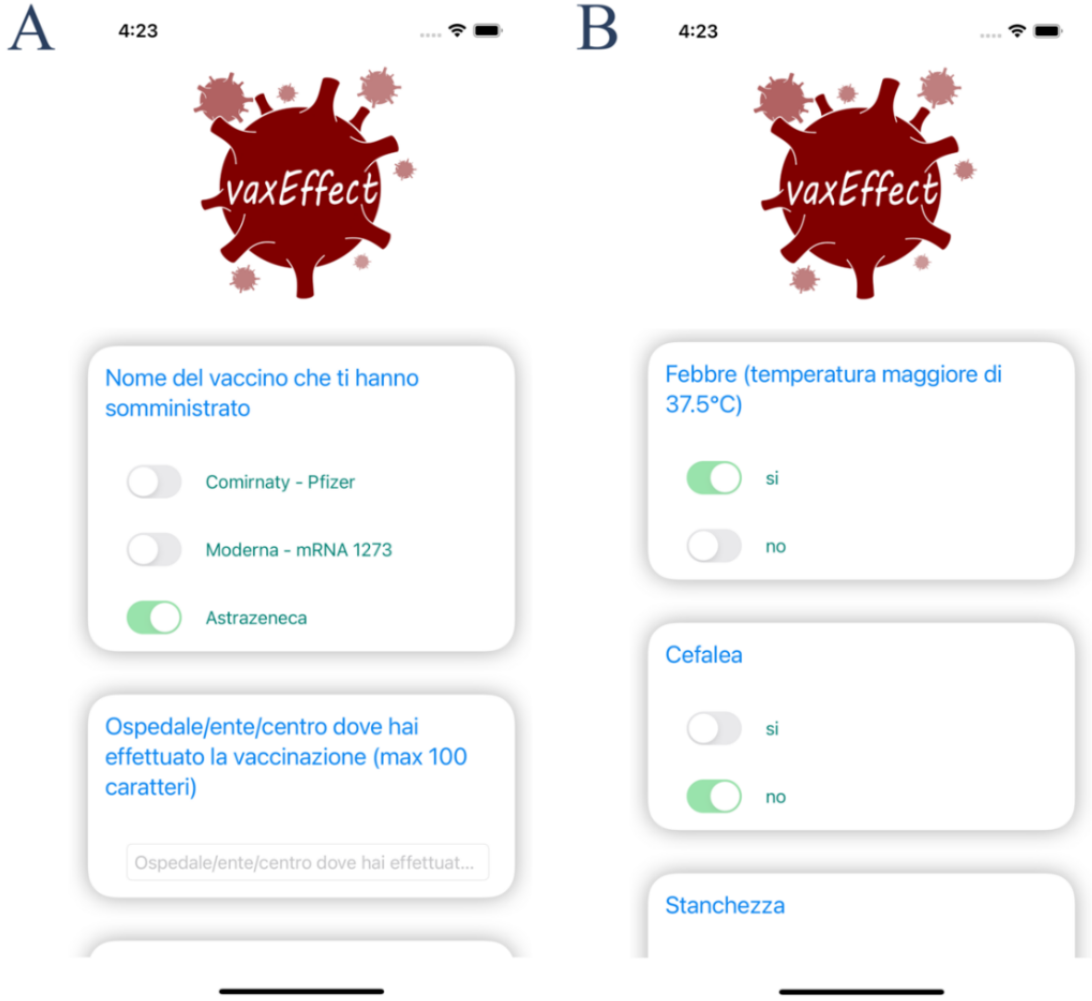


Figure 4. Mobile app at subsequent startup. (A) Vaccine data. After each vaccination, the app asks the user to submit vaccination data. In this panel, an example of the data that the user needs to submit at this stage is provided by asking the name of the administered vaccine (choosing among Comirnaty, moderna-mRNA1273, or Vaxzevria) and the hospital or institution where the vaccine has been administered. (B) ADRs. Seven days after each vaccination, the user will be asked to submit specific ADRs among those reported in the "Data collected by the app" section. In this panel, an example of the data that the user needs to submit at this stage is provided by asking whether fever, headache, or tiredness has occurred after vaccination. ADR: adverse drug reaction.



Application Server

The application server was developed in Python. Its main purpose is to collect data from the mobile app and store them in the database. The application server is also responsible for generating the UUID, to pair it with the FCM token, and to send push notification requests to the FCM server. Indeed, according to the data provided by the patient, after 7 days from the date of vaccination, the application server sends a push notification request to the FCM server in order to inform the patient that the ADR data need to be collected. The notification messages delivered to the mobile app from the FCM server open when the patient taps on the notification.

Data Collected by the App

In addition to demographic data, the medical history questionnaire that respondents had to complete required them to indicate whether they had been diagnosed with one or more chronic diseases among those suggested by the app: dyslipidemia, hypertension, autoimmune diseases, asthma, cardiovascular diseases, tumor, liver diseases, diabetes, bronchitis, or kidney disease.

In addition, vaxEffect@UniMiB asks users to register the occurrence of specific ADRs among those reported as a COVID-19 vaccine side effect by the Centers for Disease Control and Prevention (CDC) and WHO: fever, headache, injection site pain, tiredness, muscular pain, swollen lymph nodes, joint pain, paresthesia, dizziness, sleepiness, nausea, and abdominal pain. At the same stage, vaxEffect@UniMiB collects information about the vaccine administrated to the subject, such as its kind: Vaxzevria (AstraZeneca), Comirnaty (Pfizer/BioNTech), moderna-mRNA1273 (Moderna), or Janssen (Johnson & Johnson).

Data Analysis

To avoid cases of intentional or unintentional false reports by users, we checked the consistency of the submitted data at different time points, discarding nonconsistent data. Due the free access to the app and the data themselves, it was not possible to automatically detect fraudulent data reporting, but during a manual inspection, no fraudulent report was found.

For each user, we labeled the registered data, defining whether they belonged to the first, second, or third vaccination dose. By

keeping separate the 3 sets of records, we analyzed in each the number of ADRs registered by each user by separating into 2 groups subjects who reported more than 3 postvaccination symptoms and those who reported fewer than 3 symptoms. These 2 groups of subjects were compared in terms of sex and age by using the chi-square test [63].

Through radar plots, we analyzed the kinds of ADRs reported by each user after every vaccine dose by keeping subjects separated in terms of gender and age.

Using a Sankey diagram [64], we followed the ADRs reported by users who submitted their data after all 3 vaccine doses, classifying who registered 0, 1, 2, 3, or more than 3 postvaccination symptoms.

Results

Characterization of vaxEffect@UniMiB Users

The total number of people vaccinated during the first wave of vaccine administration (March 2021) conducted at the Bicocca hub was 3712 (a third of the entire personnel of UniMiB and PoliMi). This partial coverage of the personnel can be ascribed to the fact that health care workers received the vaccine before the launching of the app, with the effect that not everyone may have necessarily participated in the survey presented in this work.

During this wave, vaxEffect@UniMiB received a total of 2733 medical history questionnaires from individuals between the ages of 19 and 80 years. In particular, as shown in the Medical History column of Tables 1-3, a greater involvement of UniMiB users compared to PoliMi users resulted in more usage of the app (n=1676, 61.3%, vs n=1057, 38.7%) at UniMiB due to how vaccinations at the UniMiB hub were organized. Indeed, all the UniMiB employees were allowed to get vaccinated at the hub. On the contrary, only part of PoliMi employees were allowed to get vaccinated at this hub.

vaxEffect@UniMiB allows characterizing the vaccinated population by the kind of administered vaccine. It is possible to observe that the administered vaccines were Vaxzevria (n=2711, 62.8%), Comirnaty (n=1253, 29.0%), moderna-mRNA1273 (n=349, 8.1%), and Janssen (n=5, 0.1%). The collected data made it possible to follow the proportion of each vaccine kind through the 3 doses (Figure 5).

Table 1. Characterization of vaxEffect@UniMiB users for the institution stratification.[a]

Institution	Medical history (N=2733), n (%)	Dose 1		Dose 2		Dose 3	
		Admin.[b] (N=2360), n (%)	ADR[c] (N=2226), n (%)	Admin. (N=1779), n (%)	ADR (N=1610), n (%)	Admin. (N=179), n (%)	ADR (N=169), n (%)
UniMiB[d]	1676 (61.3)	1383 (58.6)	1332 (59.8)	1252 (70.4)	1138 (70.7)	116 (64.8)	111 (65.7)
PoliMi[e]	1057 (38.7)	977 (41.1)	894 (40.2)	527 (29.6)	472 (29.3)	63 (35.2)	58 (34.3)

[a]Each cell reports the absolute and the relative frequency computed for the sample size N reported in the column headings.

[b]Admin.: individuals who submitted their vaccination data.

[c]ADR: adverse drug reaction (individuals who registered their ADRs 7 days postvaccination).

[d]UniMiB: University of Milano-Bicocca.

[e]PoliMi: Politecnico of Milan.

Table 2. Characterization of vaxEffect@UniMiB users for the gender stratification.[a]

Gender	Medical history (N=2733), n (%)	Dose 1		Dose 2		Dose 3	
		Admin.[b] (N=2360), n (%)	ADR[c] (N=2226), n (%)	Admin. (N=1779), n (%)	ADR (N=1610), n (%)	Admin. (N=179), n (%)	ADR (N=169), n (%)
Male	1224 (44.8)	1075 (45.6)	998 (44.8)	713 (40.1)	645 (40.1)	68 (38.0)	65 (38.5)
Female	1509 (55.2)	1285 (54.4)	1228 (55.2)	1066 (59.9)	965 (59.9)	111 (62.0)	104 (61.5)

[a]Each cell reports the absolute and the relative frequency computed for the sample size N reported in the column headings.

[b]Admin.: individuals who submitted their vaccination data.

[c]ADR: adverse drug reaction (individuals who registered their ADRs 7 days postvaccination).

Table 3. Characterization of vaxEffect@UniMiB users for the age stratification.[a]

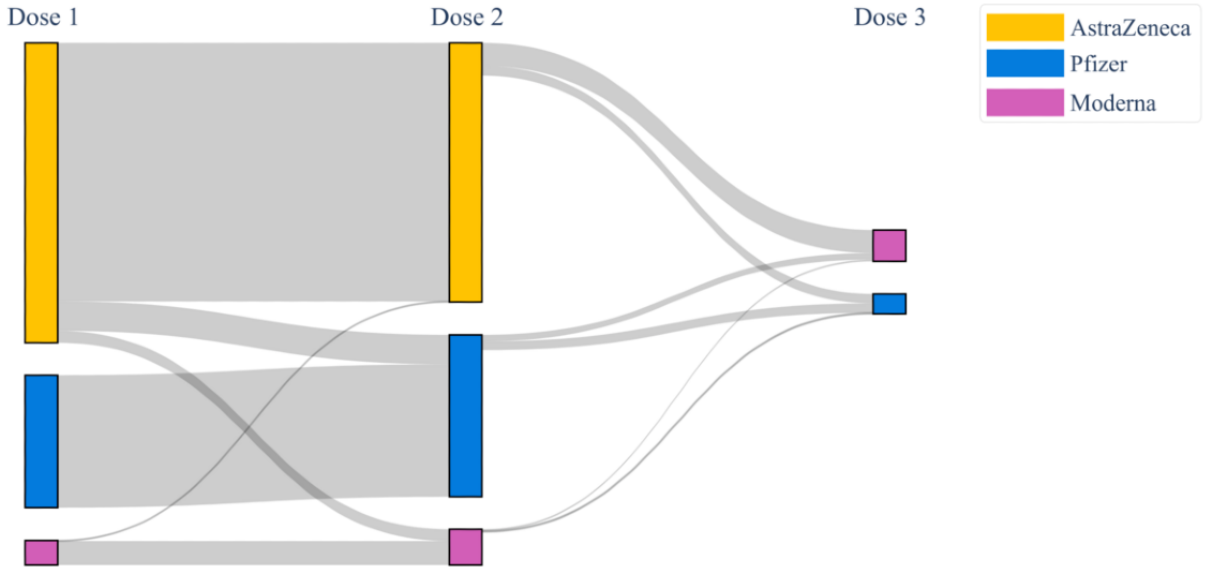
Age (years)	Medical history (N=2733), n (%)	Dose 1		Dose 2		Dose 3	
		Admin.[b] (N=2360), n (%)	ADR[c] (N=2226), n (%)	Admin. (N=1779), n (%)	ADR (N=1610), n (%)	Admin. (N=179), n (%)	ADR (N=169), n (%)
<30	1301 (47.6)	1024 (43.4)	962 (43.2)	955 (53.7)	861 (53.5)	64 (35.8)	59 (34.9)
30-55	1109 (40.6)	1043 (44.2)	983 (44.2)	634 (35.6)	575 (35.7)	80 (44.7)	76 (45.0)
>55	323 (11.8)	293 (12.4)	281 (12.6)	190 (10.7)	174 (10.8)	35 (19.5)	34 (20.1)

[a]Each cell reports the absolute and the relative frequency computed for the sample size N reported in the column headings.

[b]Admin.: individuals who submitted their vaccination data.

[c]ADR: adverse drug reaction (individuals who registered their ADRs 7 days postvaccination).

Figure 5. Kinds of vaccines administered to the same subject through the 3 vaccination campaigns.



Stratification of subjects by gender revealed a slightly greater proportion of female (n=1509, 55.2%) compared to male (n=1224, 44.8%) subjects. When stratified by age, younger (<30 years) and middle-aged (30-55 years) subjects represented the most active users, together constituting 88.2% (n=2410) of the total participants. On the contrary, older users (>55 years) represented just 11.8% (n=323) of the total set of survey respondents. Age and sex distributions of users employed in UniMiB were in line with those in PoliMi.

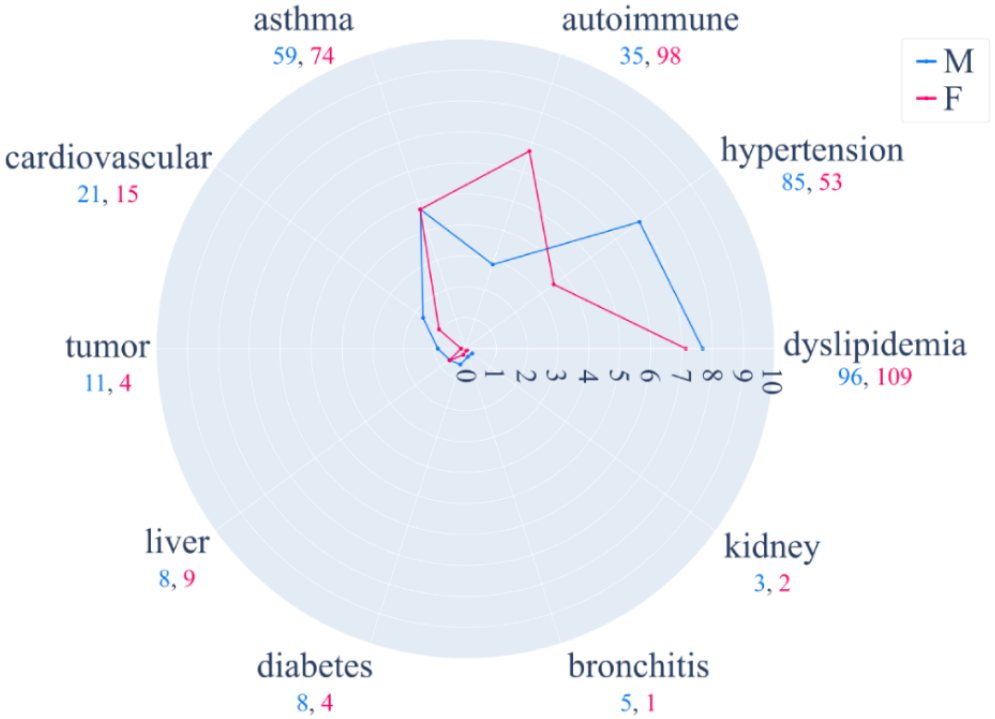
Once the medical history questionnaire was completed, 2360 (86.3%) and 1779 (65.1%) of all survey participants continued to use vaxEffect@UniMiB, reporting their vaccination data of, respectively, the first and the second dose campaign. Subsequently, 2226 (94.3%) and 1610 (90.5%) users who submitted their vaccination data after, respectively, the first and the second doses also reported their ADRs to the vaccination.

In response to the first vaccine dose, we collected information about vaccination and ADRs for 2360 and 2226 subjects (see the Dose 1 column in Tables 1-3), which represented, respectively, 63.6% and 59.9% of the 3712 individuals who were vaccinated during the first dose campaign. After the second vaccine dose, we collected information about vaccination and ADRs for 1779 and 1610 subjects (see the Dose 2 column in Tables 1-3), which represented, respectively, 47.8% and 43.3% of the 3718 individuals who were vaccinated during the second dose campaign.

The Dose 3 column in Tables 1-3 reports the same information relative to the administration of the third vaccine dose. Users who went further in reporting their data were considerably lower in number compared to the first two dose campaigns, since 179 users spontaneously continued to use vaxEffect@UniMiB to report their vaccination data and 169 of them also submitted their postvaccination ADRs. In this vaccination campaign, differently from the previous 2 ones, UniMiB was no longer the vaccination hub.

Stratifying the subjects who participated in the survey by gender (Figure 6) the prevalence of chronic diseases was overall fairly low since a maximum value of 0.078 was registered. Exploring the inserted data, the prevalence of the registered chronic diseases was higher in male than in female users, except for asthma, liver disorders, and autoimmune diseases. Moreover, focusing on cases where the gap between males and females was wider, we found that male subjects suffer more from hypertension, whereas females are more prone to autoimmune diseases.

Figure 6. Prevalence expressed in percentage of chronic diseases among male (M) and female (F) users of vaxEffect@UniMiB. The blue and pink lines depict the percentage prevalence of each single chronic disease between, respectively, male and female subjects. Below each label, the corresponding absolute count in male and female users is reported colored according to the color scale on the right.



Characterization of ADRs After Each Vaccination Dose

As shown in Figure 7, independently of the type of vaccine and the dose administered, the response of subjects in terms of the number of reported ADRs to vaccination depended on gender and age. Indeed, females were more prone to a worse response to the vaccine, given the constant higher proportion of females compared to males who registered more than 3 symptoms/week postvaccination. In terms of age, the prevalence of ADRs to vaccination turned out to be worse in younger and middle-aged respondents compared to older users.

Moving on to a more in-depth analysis of the kinds of symptoms reported after each vaccination dose, the most common vaccine ADRs were fever, headache, injection site pain, tiredness, muscular pain, joint pain, and sleepiness (Figure 8).

The types of ADRs to vaccine administration were dependent on both the gender and age of the participants. When analyzing the response of males and females separately, a progressive reduction in ADR prevalence through the 3 vaccine doses emerged for all the 3 age groups. However, we noted an exacerbation after the third vaccination dose for some of them. Specifically, among female subjects, more cases of fever, headache, injection site pain, and muscular pain were reported by young users; more cases of injection site pain, muscular pain, joint pain, and sleepiness by middle-aged users; and more cases of headache and injection site pain by older users. Instead, among male subjects, more cases of headache, injection site pain, tiredness, muscular pain, and joint pain were reported by young users; more cases of fever, headache, injection site pain, muscular pain, and joint pain by middle-aged users; and more cases of injection site pain by older users. Comparing males and females within the same vaccination dose, the same behavior as seen in Figure 7 emerged since females complained about more ADRs after each vaccine administration compared to male participants.

Figure 7. Number of ADRs registered by users after each vaccination dose. Subjects separately stratified by gender (A) and age (B) were grouped according to the number of postvaccination symptoms (>=3 Sympt. and <3 Sympt.). On the top of each plot, the result of the chi-square test and the corresponding *P* value are reported. ADR: adverse drug reaction.

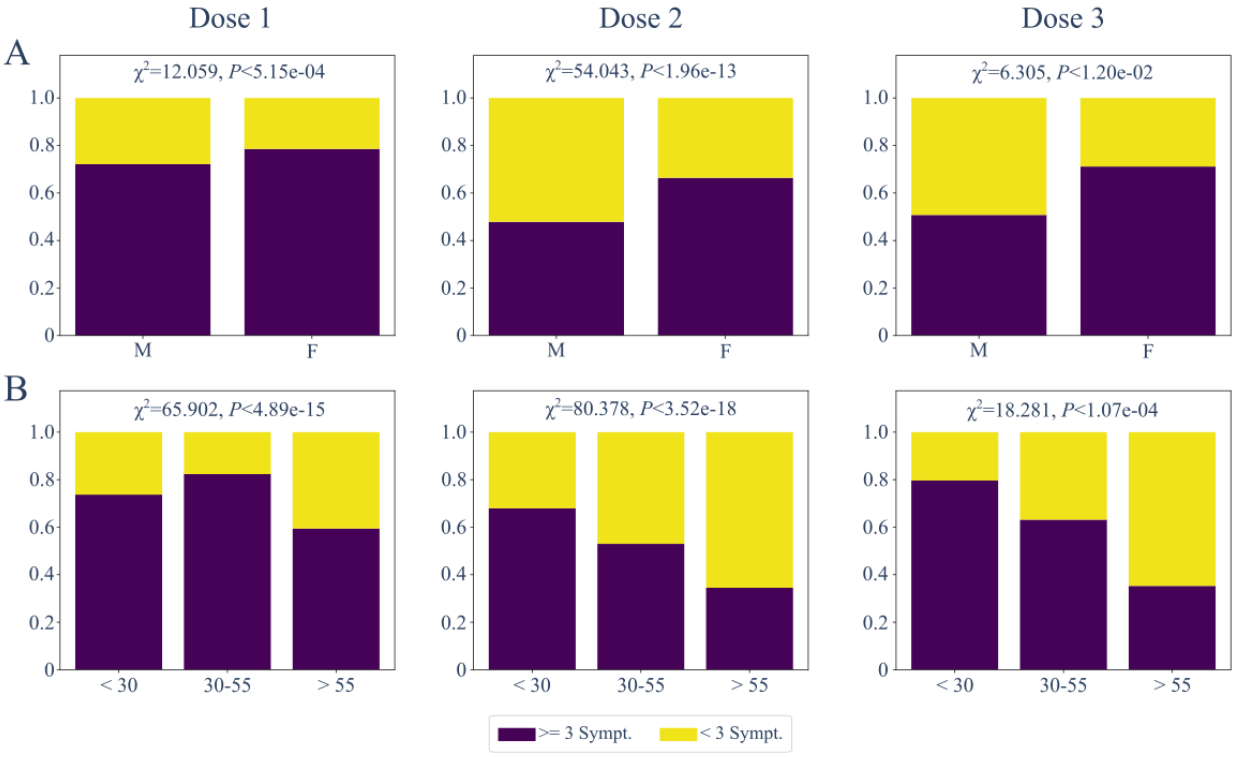
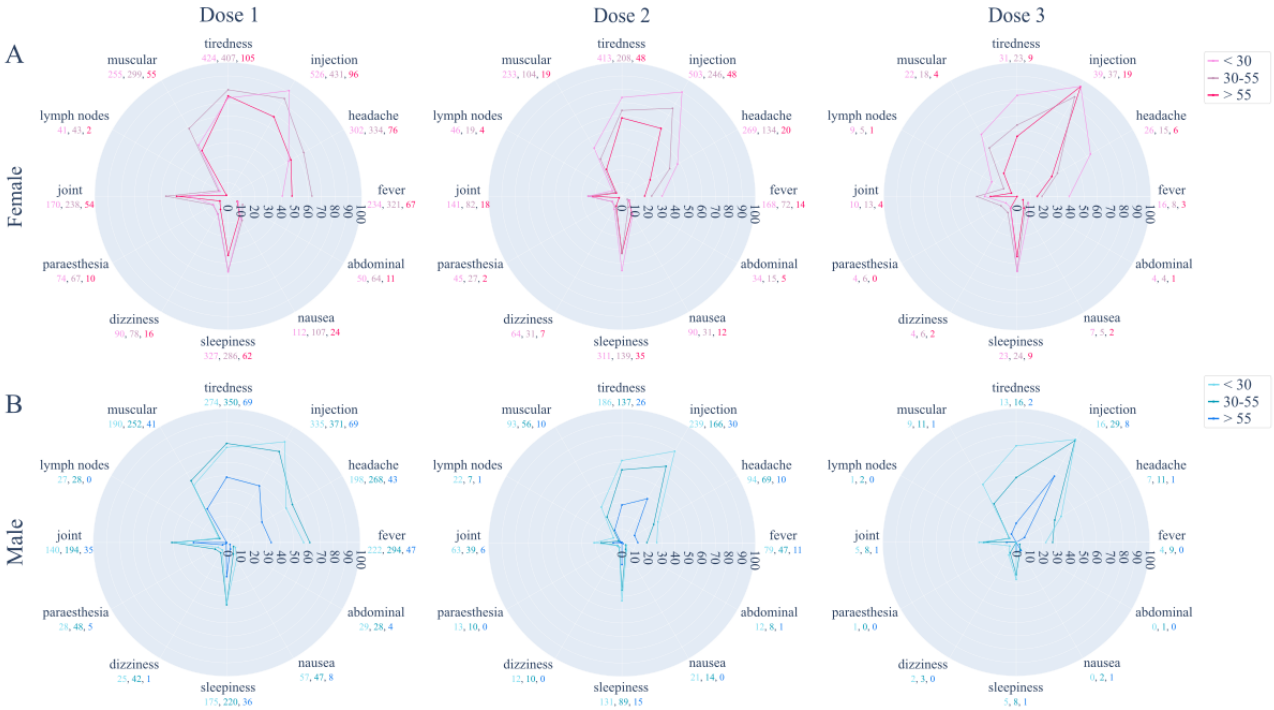


Figure 8. Kinds of ADRs registered by users after each vaccination dose. (A) Light-, middle-, and dark-pink lines denote the percentage prevalence of ADRs reported, respectively, by younger (labeled as <30 years), middle-aged (labeled as 30-55 years), and older (labeled as >55 years) female users. (B) Light-, middle-, and dark-blue lines denote the percentage prevalence of ADRs reported, respectively, by younger (labeled as <30 years), middle-aged (labeled as 30-55 years), and older (labeled as >55 years) male users. Below each label, the corresponding absolute count in younger, middle-aged, and older users is reported for females and males, colored according to the color scale on the right. ADR: adverse drug reaction.

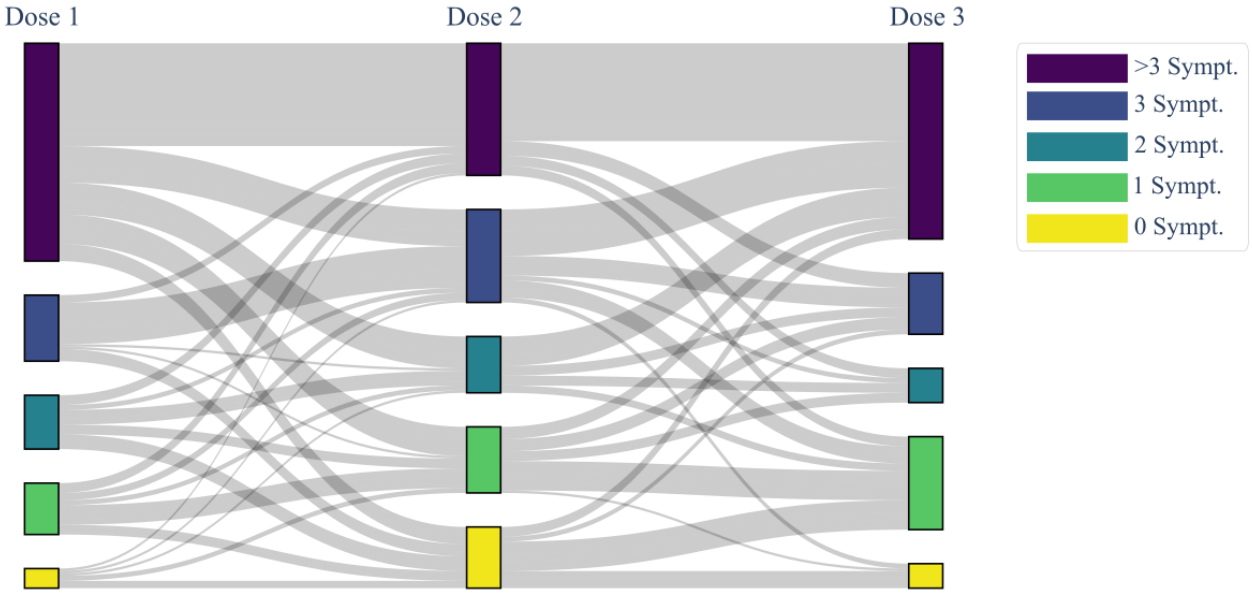


Longitudinal Analysis of ADRs Through the 3 Vaccine Doses

The innovative feature that distinguishes vaxEffect@UniMiB from the existing mobile apps aimed at reporting ADRs is the ability to perform a longitudinal analysis to monitor the same subject over time in terms of the reported ADRs after each vaccine administration, while guaranteeing anonymity of the submitted data.

In the overall cohort of the survey presented in this work, 167 users regularly submitted their ADRs after each of the 3 doses, enabling us to detect how the response of individuals changes depending on the number of ADRs registered after each dose (Figure 9).

Figure 9. Longitudinal analysis of ADRs through the 3 vaccine doses. Each level identifies each of the 3 vaccine doses, respectively, labeled as Dose 1, Dose 2, and Dose 3. Each node identifies the 5 classes of symptom abundance, respectively, labeled as 0 Sympt., 1 Sympt., 2 Sympt., 3 Sympt., and >3 Sympt. and colored according to the legend on the right. The width of the links is proportional to the amount of subjects who experienced a given number of ADRs between Dose 1 and Dose 2 and between Dose 2 and Dose 3. ADR: adverse drug reaction.



Discussion

Principal Results

We proposed a new mobile app called vaxEffect@UniMiB, able to perform anonymized longitudinal studies, that, combined with the health care workers' sponsorships, fostered the spontaneous citizenship participation in COVID-19 vaccination ADR data collection. In this work, over 3712 subjects were vaccinated during the first dose at the Bicocca hub, and we received a total of 2733 (73.6%) reports.

It should be noted that health care workers, receiving the vaccine before the launching of the app, may not have necessarily participated in the survey presented in this work. Moreover, due to deputed authority emergency plan, the hub invited all UniMiB employees and only part of PoliMi ones, the others being distributed to different hubs. This aspect further affected the number of reports received over the total number of subjects invited to the vaccination campaign.

Thanks to vaxEffect@UniMiB, we collected vaccination and ADR data relative to the first vaccine dose for 2226 (60.0%) of the 3712 individuals vaccinated during the first dose campaign, to the second dose for 1610 (43.3%) of the 3718 individuals vaccinated during the second dose campaign, and, in a nonsponsored fashion, to the third dose for 169 (4.6%) individuals.

Although a direct comparison of our reporting rate against traditional ones is not possible, a reporting rate of 120 questionnaires submitted to the National Pharmacovigilance Network for any 100,000 doses administered was reported, according to the Annual Report of Aifa (Agenzia italiana del farmaco) related to the surveillance of COVID-19 vaccine safety during December 27, 2020-December 26, 2021.

Following a review of the currently available mobile apps for monitoring ADR reporting for medical devices, drugs, or vaccines, our app revealed to be one step ahead. Specifically, vaxEffect@UniMiB is the first successful attempt in tracking the same user over time and identifying putative temporal patterns in an anonymous fashion. With this ability, we were able to analyze in this study the trend that emerged through the 3 vaccine doses administered to the same user. The first vaccination resulted in the strongest response among individuals, given the high prevalence of subjects with 3 or more than 3 symptoms. As already evidenced by Haas et al [65], part of this high response to the vaccine could be due to a misattribution of commonly experienced and nonspecific symptoms, such as headache and tiredness, as specific reactions due to the vaccine administration, instead of a condition of anxiety and worry making people hyperalert to the occurrence of any possible adverse events. Among the 3 administered doses, the second one was the least symptomatic since a general decreased trend in the adverse events from the first to the second dose emerged.

Finally, between the second and the third dose, there was a predominantly increased trend in the number of reported symptoms. This behavior might be due to a change in the vaccine type, since after the second dose, AstraZeneca has not been administered since. However, given the low number of subjects who regularly submitted their data after all the 3 vaccine doses, we are not in a position to determine whether this increased trend may for sure be due to the change in the vaccine type that occurred between the second and the third vaccination campaign.

A series of aspects contributed to a greater involvement from people in using this app to report their ADRs to vaccination.

First, the user-friendly and interactive graphics of vaxEffect@UniMiB allowed ease of use of the app, resulting in a positive response rate after both the first and the second vaccination campaign. Although the response after the third campaign was lower compared to the previous 2 ones, it could be still considered a positive result. Indeed, during the third vaccination campaign, any reminder notification was sent to users, denoting a completely spontaneous participation of subjects in that phase. This result indicates the effectiveness of vaxEffect@UniMiB as an ADR-reporting system, although the active contribution of an institutional sponsorship is necessary.

Another aspect that contributed to the increase in the usage of vaxEffect@UniMiB was its availability for both Android and iOS devices, which made it reachable to as many users as possible.

Even more importantly, the protection of the submitted information assured users that high standards of confidentiality were maintained during data reporting. The architecture underlying vaxEffect@UniMiB ensured the anonymity of all the survey participants and also the privacy of their data, minimizing the illegitimate acquisition of information in the case of loss, theft, or fraudulent use of the device. These actions allowed guaranteeing proper use of the outcomes of the research, while preserving the possibility of tracking the same subject over time.

Limitations

Although these features allowed for a positive response from the public, we are also aware that following an even stronger sponsorship of the vaccination campaign, the 373 users who just compiled the medical questionnaire without then reporting any other data would continue using the app. Likewise, increased support in the dissemination of the app could encourage who partially submitted data for only 1 or 2 vaccine doses to complete the entire survey.

We intend to also clarify that since the study presented in this work was a pilot study conducted on a selected voluntary population, we cannot exclude that our statistical measures could be affected by volunteer bias. This bias can occur at all stages of the study, and differences between volunteers and the target population are not restricted to sociodemographic factors but can include attitudes toward the study and institutions involved. However, the likelihood of volunteer bias increases as the refusal rate to volunteer increases. We observed a high rate of participation and a low rate of attraction until the third vaccination wave (when the Bicocca hub was not more in place and the in-person stimulus ended), probably due to the endorsement of participants' anonymity and confidentiality.

The impact of this bias is unknown, but the observed sample should be quite younger and have a higher level of education and income than the target population, showing a higher rate of participation. These characteristics have been associated with greater extension and severity, so we cannot exclude that we overestimated the prevalence of symptoms [66].

This work being a pilot study, we chose to restrict the analyses to the respondents explicitly imputable to the hub itself. Consequently, data entry was strictly controlled since only ADRs to the administered vaccination were asked to be submitted. In this way, we were sure that all the reported information was not false-true ADRs due to other medical conditions of the user. However, as a suggestion for the future, if vaxEffect@UniMiB will be used in a different and more open context, it will be necessary to integrate a preprocessing phase to tame the possible distortions induced by the noncontrolled environment in which respondents report information.

Conclusion

The results of this study showed that it is possible to leverage active citizenship engagement to increase the ADR-reporting rate using an agile and pervasive tool, such as a mobile app.

The combination between the synergy of the physician who proposed the app and the flexibility of vaxEffect@UniMiB seems to be an excellent tool to ensure high response rates in situations in which a priori information is scarce and decisions need to be made in a short time.

In addition to the promising results so far obtained, this work highlighted some practices that need to be carried out to foster approach efficacity, such as conspicuous support in the dissemination of the app to encourage more subjects to report their ADRs and steady user engagement to sustain data submission through the entire vaccination treatment when only partial information is reported.

Acknowledgments

AA, AZ, and DP performed conceptualization; AA and DP, data curation; MD, AA, and DP, formal analysis; AZ and DP, funding acquisition; MB, MEP, and MAR, investigation; AA, AZ, and DP, methodology; AZ and DP, project administration; MB, MEP, MAR, and DP, resources; AA and DP, software; AZ and DP, supervision; MD and DP, visualization; MD, AA, AZ, and DP, writing—original draft; and MD, AA, AZ, DP, and MAR, writing—review and editing.

Conflicts of Interest

None declared.

References

1. World Health Organization. The Importance of Pharmacovigilance. Geneva: World Health Organization; 2002.
2. Pappa D, Stergioulas LK. Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions. Int J Data Sci Anal 2019 Feb 12;8(2):113-135 [FREE Full text] [doi: 10.1007/s41060-019-00175-3]
3. Fernandez S, Lenoir C, Samer C, Rollason V. Drug interactions with apixaban: a systematic review of the literature and an analysis of VigiBase, the World Health Organization database of spontaneous safety reports. Pharmacol Res Perspect 2020 Oct 02;8(5):e00647 [FREE Full text] [doi: 10.1002/prp2.647] [Medline: 32881416]
4. Moro PL, Woo EJ, Marquez P, Cano M. Monitoring the safety of high-dose, trivalent inactivated influenza vaccine in the vaccine adverse event reporting system (VAERS), 2011 - 2019. Vaccine 2020 Aug 18;38(37):5923-5926. [doi: 10.1016/j.vaccine.2020.07.007] [Medline: 32709434]
5. Logan JK, Wickramaratne Senarath Yapa S, Harinstein L, Saluja B, Muñoz M, Sahajwalla C, et al. Drug interaction between febuxostat and thiopurine antimetabolites: a review of the FDA adverse event reporting system and medical literature. Pharmacotherapy 2020 Feb 20;40(2):125-132. [doi: 10.1002/phar.2362] [Medline: 31885095]
6. Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS). Vaccine 2015 Aug 26;33(36):4398-4405 [FREE Full text] [doi: 10.1016/j.vaccine.2015.07.035] [Medline: 26209838]
7. Fang H, Lin X, Zhang J, Hong Z, Sugiyama K, Nozaki T, et al. Multifaceted interventions for improving spontaneous reporting of adverse drug reactions in a general hospital in China. BMC Pharmacol Toxicol 2017 Jun 26;18(1):49 [FREE Full text] [doi: 10.1186/s40360-017-0159-0] [Medline: 28651624]
8. Qing-ping S, Xiao-dong J, Feng D, Yan L, Mei-ling Y, Jin-xiu Z, et al. Consequences, measurement, and evaluation of the costs associated with adverse drug reactions among hospitalized patients in China. BMC Health Serv Res 2014 Feb 17;14(1):73 [FREE Full text] [doi: 10.1186/1472-6963-14-73] [Medline: 24533894]
9. Ng WL, Lim KS, Hariraj V, Lee SC, Wo WK, Ramli A, et al. Incidence of allopurinol-induced severe cutaneous adverse drug reaction in Malaysia. Br J Clin Pharmacol 2022 Aug;88(8):3782-3788 [FREE Full text] [doi: 10.1111/bcp.15327] [Medline: 35318720]
10. Salathé M. Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. J Infect Dis 2016 Dec 01;214(suppl_4):S399-S403 [FREE Full text] [doi: 10.1093/infdis/jiw281] [Medline: 28830106]
11. Psihogios A, Bota AB, Mithani SS, Greyson D, Zhu DT, Fung SG, et al. A scoping review of active, participant-centred, digital adverse events following immunization (AEFI) surveillance: a Canadian immunization research network study. Vaccine 2022 Jul 29;40(31):4065-4080 [FREE Full text] [doi: 10.1016/j.vaccine.2022.04.103] [Medline: 35680501]
12. Heininger U, Holm K, Caplanusi I, Bailey SR, Abdoellah SA, Arellano F. Guide to active vaccine safety surveillance: report of CIOMS working group on vaccine safety-executive summary. Vaccine 2017;35(32):3917-3921 [FREE Full text] [doi: 10.56759/hnuw8440]
13. Cashman P, Macartney K, Khandaker G, King C, Gold M, Durrheim DN. Participant-centred active surveillance of adverse events following immunisation: a narrative review. Int Health 2017 May 01;9(3):164-176 [FREE Full text] [doi: 10.1093/inthealth/ihx019] [Medline: 28582563]
14. Liu F, Jagannatha A, Yu H. Towards drug safety surveillance and pharmacovigilance: current progress in detecting medication and adverse drug events from electronic health records. Drug Saf 2019 Jan 16;42(1):95-97 [FREE Full text] [doi: 10.1007/s40264-018-0766-8] [Medline: 30649734]
15. Palleria C, Leporini C, Chimirri S, Marrazzo G, Sacchetta S, Bruno L, et al. Limitations and obstacles of the spontaneous adverse drugs reactions reporting: two “challenging” case reports. Journal of Pharmacology and Pharmacotherapeutics 2013 Dec 01;4(1_suppl):S66-S72. [doi: 10.4103/0976-500x.120955]
16. Salter S, Singh G, Nissen L, Trentino K, Murray K, Lee K, et al. Active vaccine safety surveillance of seasonal influenza vaccination via a scalable, integrated system in Western Australian pharmacies: a prospective cohort study. BMJ Open 2021 Jun 08;11(6):e048109 [FREE Full text] [doi: 10.1136/bmjopen-2020-048109] [Medline: 34103321]
17. Cashman P, Moberley S, Dalton C, Stephenson J, Elvidge E, Butler M, et al. Vaxtracker: active on-line surveillance for adverse events following inactivated influenza vaccine in children. Vaccine 2014 Sep 22;32(42):5503-5508 [FREE Full text] [doi: 10.1016/j.vaccine.2014.07.061] [Medline: 25077424]
18. Gold MS, Lincoln G, Cashman P, Braunack-Mayer A, Stocks N. Efficacy of m-Health for the detection of adverse events following immunization: the stimulated telephone assisted rapid safety surveillance (STARSS) randomised control trial. Vaccine 2021 Jan 08;39(2):332-342. [doi: 10.1016/j.vaccine.2020.11.056] [Medline: 33279317]
19. Regan AK, Blyth CC, Tracey L, Mak DB, Richmond PC, Effler PV. Comparison of text-messaging to voice telephone interviews for active surveillance of adverse events following immunisation. Vaccine 2015 Jul 17;33(31):3689-3694. [doi: 10.1016/j.vaccine.2015.06.022] [Medline: 26079616]

20. Tracey LE, Regan AK, Mak DB, Effler PV. Adverse events following influenza immunization reported by healthcare personnel using active surveillance based on text messages. Infect Control Hosp Epidemiol 2015 May 05;36(5):608-610. [doi: 10.1017/ice.2015.16] [Medline: 25652211]
21. van Balveren-Slingerland L, Kant A, Härmark L. Web-based intensive monitoring of adverse events following influenza vaccination in general practice. Vaccine 2015 May 05;33(19):2283-2288. [doi: 10.1016/j.vaccine.2015.03.014] [Medline: 25797367]
22. Newes-Adeyi G, Greece J, Bozeman S, Walker DK, Lewis F, Gidudu J. Active surveillance for influenza vaccine adverse events: the integrated vaccine surveillance system. Vaccine 2012 Feb 01;30(6):1050-1055. [doi: 10.1016/j.vaccine.2011.12.041] [Medline: 22200501]
23. O'Connor MN, O'Sullivan D, Gallagher PF, Eustace J, Byrne S, O'Mahony D. Prevention of hospital-acquired adverse drug reactions in older people using screening tool of older persons' prescriptions and screening tool to alert to right treatment criteria: a cluster randomized controlled trial. J Am Geriatr Soc 2016 Aug;64(8):1558-1566 [FREE Full text] [doi: 10.1111/jgs.14312] [Medline: 27365262]
24. Herdeiro MT, Ribeiro-Vaz I, Ferreira M, Polónia J, Falcão A, Figueiras A. Workshop- and telephone-based interventions to improve adverse drug reaction reporting. Drug Saf 2012 Dec 13;35(8):655-665. [doi: 10.1007/BF03261962]
25. Egberts KM, Gerlach M, Correll CU, Plener PL, Malzahn U, Heuschmann P, et al. Serious adverse drug reactions in children and adolescents treated on- and off-label with antidepressants and antipsychotics in clinical practice. Pharmacopsychiatry 2022 Sep 07;55(5):255-265 [FREE Full text] [doi: 10.1055/a-1716-1856] [Medline: 35130562]
26. Potlog Shchory M, Goldstein LH, Arcavi L, Shihmanter R, Berkovitch M, Levy A. Increasing adverse drug reaction reporting: how can we do better? PLoS One 2020 Aug 13;15(8):e0235591 [FREE Full text] [doi: 10.1371/journal.pone.0235591] [Medline: 32790671]
27. Nolte H, Bernstein DI, Sussman GL, Svanholm Fogh B, Lu S, Husøy B, et al. Impact of adverse event solicitation on the safety profile of SQ house dust mite sublingual immunotherapy tablet. J Allergy Clin Immunol Pract 2018;6(6):2081-2086.e1 [FREE Full text] [doi: 10.1016/j.jaip.2018.01.037] [Medline: 29432959]
28. Coates M, Spanos M, Parmar P, Chandrasekhar T, Sikich L. A review of methods for monitoring adverse events in pediatric psychopharmacology clinical trials. Drug Saf 2018 May 9;41(5):465-471 [FREE Full text] [doi: 10.1007/s40264-017-0633-z] [Medline: 29318515]
29. Lindell-Osuagwu L, Sepponen K, Farooqui S, Kokki H, Hämeen-Anttila K, Vainio K. Parental reporting of adverse drug events and other drug-related problems in children in Finland. Eur J Clin Pharmacol 2013 Apr 24;69(4):985-994. [doi: 10.1007/s00228-012-1426-z] [Medline: 23093040]
30. Verden A, Dimbil M, Kyle R, Overstreet B, Hoffman KB. Analysis of spontaneous postmarket case reports submitted to the FDA regarding thromboembolic adverse events and JAK inhibitors. Drug Saf 2018 Apr;41(4):357-361 [FREE Full text] [doi: 10.1007/s40264-017-0622-2] [Medline: 29196988]
31. Abadie D, Chebane L, Bert M, Durrieu G, Montastruc JL. Addendum to: Online reporting of adverse drug reactions: a study from a French regional pharmacovigilance center. Thérapie 69(5): 395-400 (2014) DOI:10.2515/therapie/2014035. Therapie 2015 Oct 16:Online ahead of print. [doi: 10.2515/therapie/2015058] [Medline: 26475752]
32. Huang CY, Hsieh YW, Lin WL, Wu HC. PCN1 the evaluation of ADR online reporting of radiopaque agents in a university hospital in TAIWAN. Value Health 2010 Nov;13(7):A510. [doi: 10.1016/s1098-3015(11)73094-7]
33. Ribeiro-Vaz I, Santos C, da Costa-Pereira A, Cruz-Correia R. Promoting spontaneous adverse drug reaction reporting in hospitals using a hyperlink to the online reporting form: an ecological study in Portugal. Drug Saf 2012 May 01;35(5):387-394. [doi: 10.2165/11597190-000000000-00000] [Medline: 22468615]
34. Jeon M, Kim J, Oh CE, Lee JY. Adverse events following immunization associated with the first and second doses of the ChAdOx1 nCoV-19 vaccine among healthcare workers in Korea. Vaccines (Basel) 2021 Sep 28;9(10):1096 [FREE Full text] [doi: 10.3390/vaccines9101096] [Medline: 34696204]
35. Song H, Pei X, Liu Z, Shen C, Sun J, Liu Y, et al. Pharmacovigilance in China: evolution and future challenges. Br J Clin Pharmacol 2022 Feb 14:Online ahead of print. [doi: 10.1111/bcp.15277] [Medline: 35165914]
36. Lander AR, Blicher TM, Jimenez-Solem E, Jespersen M, Kampmann JP, Christensen HR. Introducing an adverse drug event manager. Eur J Hosp Pharm 2013 Feb 20;20(2):78-81. [doi: 10.1136/ejhpharm-2012-000171]
37. Lee SB, Schepers GP, Goldberg KL. Electronic adverse-drug-reaction-reporting program. Am J Health Syst Pharm 2004 Jun 15;61(12):1230, 1232-1230, 1233. [doi: 10.1093/ajhp/61.12.1230] [Medline: 15259751]
38. Salvador MR, Monteiro C, Pereira L, Duarte AP. Quality of spontaneous reports of adverse drug reactions sent to a regional pharmacovigilance unit. Int J Environ Res Public Health 2022 Mar 22;19(7):3754 [FREE Full text] [doi: 10.3390/ijerph19073754] [Medline: 35409436]
39. Hawcutt DB, Mainie P, Riordan A, Smyth RL, Pirmohamed M. Reported paediatric adverse drug reactions in the UK 2000-2009. Br J Clin Pharmacol 2012 Mar;73(3):437-446 [FREE Full text] [doi: 10.1111/j.1365-2125.2011.04113.x] [Medline: 21988288]
40. Ferner RE, Stevens RJ, Anton C, Aronson JK. Spontaneous reporting to regulatory authorities of suspected adverse drug reactions to COVID-19 vaccines over time: the effect of publicity. Drug Saf 2022 Feb 22;45(2):137-144 [FREE Full text] [doi: 10.1007/s40264-021-01138-z] [Medline: 35064899]

41. Hasan SS, Bal N, Baker I, Kow CS, Khan MU. Adverse drug reaction reporting and prescribing trends of drugs for attention deficit hyperactivity disorder in primary care England, 2010-2019. J Atten Disord 2022 Feb;26(3):467-475 [FREE Full text] [doi: 10.1177/1087054721997556] [Medline: 33666114]
42. Welcome to the Yellow Card Reporting Site. URL: https://yellowcard.mhra.gov.uk/ [accessed 2022-10-07]
43. van Hunsel F, Härmark L, Pal S, Olsson S, van Grootheest K. Experiences with adverse drug reaction reporting by patients: an 11-country survey. Drug Saf 2012 Jan 01;35(1):45-60. [doi: 10.2165/11594320-000000000-00000] [Medline: 22149419]
44. Yousef NB, Yenugadhati N, Alqahtani N, Alshahrani A, Alshahrani M, Al Jeraisy M, et al. Patterns of adverse drug reactions (ADRs) in Saudi Arabia. Saudi Pharm J 2022 Jan;30(1):8-13 [FREE Full text] [doi: 10.1016/j.jsps.2021.08.014] [Medline: 35145342]
45. Wilson K, Atkinson KM, Westeinde J, Bell C, Marty K, Fergusson D, et al. An evaluation of the feasibility and usability of a proof of concept mobile app for adverse event reporting post influenza vaccination. Hum Vaccin Immunother 2016 Jul 02;12(7):1738-1748 [FREE Full text] [doi: 10.1080/21645515.2016.1152434] [Medline: 26905396]
46. Pierce CE, de Vries ST, Bodin-Parssinen S, Härmark L, Tregunno P, Lewis DJ, et al. Recommendations on the use of mobile applications for the collection and communication of pharmaceutical product safety information: lessons from IMI WEB-RADR. Drug Saf 2019 Apr 25;42(4):477-489 [FREE Full text] [doi: 10.1007/s40264-019-00813-6] [Medline: 30911975]
47. Defer G, Fedrizzi S, Chevanne D, Montastruc F, Briant AR, Parienti JJ, French VigipSEP Study Group, Société Francophone de la Sclérose en Plaques (SFSEP). Adverse drug reaction reporting using a mobile device application by persons with multiple sclerosis: a cluster randomized controlled trial. Drug Saf 2021 Feb 09;44(2):223-233 [FREE Full text] [doi: 10.1007/s40264-020-01009-z] [Medline: 33048319]
48. Ahn E, Liu N, Parekh T, Patel R, Baldacchino T, Mullavey T, et al. A mobile app and dashboard for early detection of infectious disease outbreaks: development study. JMIR Public Health Surveill 2021 Mar 09;7(3):e14837 [FREE Full text] [doi: 10.2196/14837] [Medline: 33687334]
49. Prakash J, Joshi K, Malik D, Mishra O, Sachan A, Kumar B, et al. "ADR PvPI" Android mobile app: report adverse drug reaction at any time anywhere in India. Indian J Pharmacol 2019;51(4):236. [doi: 10.4103/ijp.ijp_595_18]
50. Montastruc F, Bagheri H, Lacroix I, Damase-Michel C, Chebane L, Rousseau V, et al. Adverse drug reaction reports received through the mobile app, VigiBIP: a comparison with classical methods of reporting. Drug Saf 2018 May 21;41(5):511-514. [doi: 10.1007/s40264-017-0630-2] [Medline: 29270770]
51. de Vries ST, Denig P, Lasheras Ruiz C, Houÿez F, Wong L, Sutcliffe A, IMI Web-RADR Work Package 3b Consortium. Interest in a mobile app for two-way risk communication: a survey study among European healthcare professionals and patients. Drug Saf 2018 Jul 2;41(7):697-712 [FREE Full text] [doi: 10.1007/s40264-018-0648-0] [Medline: 29500800]
52. Bahk CY, Goshgarian M, Donahue K, Freifeld CC, Menone CM, Pierce CE, et al. Increasing patient engagement in pharmacovigilance through online community outreach and mobile reporting applications: an analysis of adverse event reporting for the Essure device in the US. Pharmaceut Med 2015;29(6):331-340 [FREE Full text] [doi: 10.1007/s40290-015-0106-6] [Medline: 26635479]
53. Unknown. Adverse drug reaction reports via VigiBIP app. Reactions Weekly 2018 Jan 13;1684(1):2-13. [doi: 10.1007/s40278-018-40614-3]
54. WEB-RADR. Mobile Apps. URL: https://web-radr.eu/mobile-apps/ [accessed 2022-10-07]
55. Netherlands Pharmacovigilance Centre Lareb. New: Side Effect App. URL: https://www.lareb.nl/en/news/new-side-effect-app [accessed 2022-10-07]
56. WEB-RADR. Med Safety App. URL: https://web-radr.eu/mobile-apps/med-safety/ [accessed 2022-10-07]
57. ADR Online. URL: https://apps.apple.com/app/adr-online/id403478954 [accessed 2022-10-07]
58. ADR Reporter. URL: https://play.google.com/store/apps/details?id=com.amu.slidingmenu [accessed 2022-10-07]
59. TMDA Adverse Reactions Reporting Tool. URL: https://play.google.com/store/apps/details?id=com.cive.HakikiDawaADR&hl=en_US&gl=US [accessed 2022-10-07]
60. Easypharm App. URL: https://www.easypharm.be/site/ [accessed 2022-10-07]
61. WEB-RADR. UAE RADR. URL: https://web-radr.eu/mobile-apps/med-safety/uae-radr/ [accessed 2022-10-07]
62. Firebase Cloud Messaging. URL: https://firebase.google.com/docs/cloud-messaging/ [accessed 2022-10-07]
63. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos Mag 2009 Apr 21;50(302):157-175. [doi: 10.1080/14786440009463897]
64. Sankey Diagram in Python. URL: https://plotly.com/python/sankey-diagram/ [accessed 2022-10-07]
65. Haas JW, Bender FL, Ballou S, Kelley JM, Wilhelm M, Miller FG, et al. Frequency of adverse events in the placebo arms of COVID-19 vaccine trials: a systematic review and meta-analysis. JAMA Netw Open 2022 Jan 04;5(1):e2143955 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.43955] [Medline: 35040967]
66. Camacho Moll ME, Salinas Martínez AM, Tovar Cisneros B, García Onofre JI, Navarrete Floriano G, Bermúdez de León M. Extension and severity of self-reported side effects of seven COVID-19 vaccines in Mexican population. Front Public Health 2022 Mar 14;10:834744 [FREE Full text] [doi: 10.3389/fpubh.2022.834744] [Medline: 35359754]

Abbreviations

ADR: adverse drug reaction
FCM: Firebase Cloud Messaging
IVSS: integrated vaccine surveillance system
PoliMi: Politecnico of Milan
SDK: software development kit
UID: user identifier
UniMiB: University of Milano-Bicocca
UUID: unique user identifier
WEB-RADR: Web–Recognising Adverse Drug Reactions
WHO: World Health Organization

Edited by A Kushniruk; submitted 20.04.22; peer-reviewed by S Rostam Niakan Kalhori, V Rajan, S Matsuda; comments to author 18.06.22; revised version received 08.07.22; accepted 02.08.22; published 04.11.22.

Please cite as:
Di Filippo M, Avellone A, Belingheri M, Paladino ME, Riva MA, Zambon A, Pescini D
A Mobile App Leveraging Citizenship Engagement to Perform Anonymized Longitudinal Studies in the Context of COVID-19 Adverse Drug Reaction Monitoring: Development and Usability Study
JMIR Hum Factors 2022;9(4):e38701
URL: https://humanfactors.jmir.org/2022/4/e38701
doi:10.2196/38701
PMID:35930561

Original Paper

Assessing the Topics and Motivating Factors Behind Human-Social Chatbot Interactions: Thematic Analysis of User Experiences

Vivian P Ta-Johnson[1], MSc, PhD; Carolynn Boatfield[1,2], BA; Xinyu Wang[1,3], BA; Esther DeCero[1,4], BA; Isabel C Krupica[1,2], BA; Sophie D Rasof[1], BA; Amelie Motzer[1]; Wiktoria M Pedryc[1]

[1]Department of Psychology, Lake Forest College, Lake Forest, IL, United States

[2]College of Health Professions, Rosalind Franklin University, North Chicago, IL, United States

[3]Department of Psychology, Columbia University, New York City, NY, United States

[4]School of Health Sciences and Public Health, Loyola University Chicago, Maywood, IL, United States

Corresponding Author:
Vivian P Ta-Johnson, MSc, PhD
Department of Psychology
Lake Forest College
555 N Sheridan Road
Lake Forest, IL, 60045
United States
Phone: 1 847 735 5258
Email: ta@lakeforest.edu

Abstract

Background: Although social chatbot usage is expected to increase as language models and artificial intelligence improve, very little is known about the dynamics of human-social chatbot interactions. Specifically, there is a paucity of research examining why human-social chatbot interactions are initiated and the topics that are discussed.

Objective: We sought to identify the motivating factors behind initiating contact with Replika, a popular social chatbot, and the topics discussed in these interactions.

Methods: A sample of Replika users completed a survey that included open-ended questions pertaining to the reasons why they initiated contact with Replika and the topics they typically discuss. Thematic analyses were then used to extract themes and subthemes regarding the motivational factors behind Replika use and the types of discussions that take place in conversations with Replika.

Results: Users initiated contact with Replika out of interest, in search of social support, and to cope with mental and physical health conditions. Users engaged in a wide variety of discussion topics with their Replika, including intellectual topics, life and work, recreation, mental health, connection, Replika, current events, and other people.

Conclusions: Given the wide range of motivational factors and discussion topics that were reported, our results imply that multifaceted support can be provided by a single social chatbot. While previous research already established that social chatbots can effectively help address mental and physical health issues, these capabilities have been dispersed across several different social chatbots instead of deriving from a single one. Our results also highlight a motivating factor of human-social chatbot usage that has received less attention than other motivating factors: interest. Users most frequently reported using Replika out of interest and sought to explore its capabilities and learn more about artificial intelligence. Thus, while developers and researchers study human-social chatbot interactions with the efficacy of the social chatbot and its targeted user base in mind, it is equally important to consider how its usage can shape public perceptions and support for social chatbots and artificial agents in general.

(JMIR Hum Factors 2022;9(4):e38876) doi:10.2196/38876

KEYWORDS

social chatbots; Replika; emotional chatbots; artificial intelligence; thematic analysis; human-chatbot interactions; chatbot; usability; interaction; human factors; motivation; topics; AI; perception; usage

Introduction

Background

With the advancement of artificial intelligence, the amount of time that people spend engaging in human-chatbot interactions will likely increase as chatbots become more ubiquitous in everyday life. This includes interactions with social chatbots—chatbots that can engender the development of companionship with human users by conversing socially and empathetically [1-3]. While social chatbot usage is on the rise [4,5], very little is known about the dynamics of these interactions, particularly about why human-social chatbot interactions are initiated and the content of such interactions [6]. In other words, what are the motivating factors behind initiating contact with a social chatbot, and what is discussed in these interactions? In this paper, we collected data from users of Replika, a popular social chatbot, to address this gap in the literature.

This investigation is important for several reasons. A prominent portion of recent chatbot research focuses on chatbot user experiences given that "the strengthening of chatbot user experiences remains a key research challenge" [7,8]. This body of work has revealed "factors contributing to positive or negative user experience…and how these aspects are impacted by chatbot design" [7]. For instance, lack of trust [9] and user dissatisfaction [10] can hinder the adoption of chatbots while affective determinants and perceived usefulness and helpfulness can improve attitudes toward chatbot usage [8]. Although this information is undoubtedly crucial for designing effective chatbots, identifying factors that contribute to a positive (or negative) user experience requires that motivating factors behind chatbot usage also be considered. This is important given that user experience is linked with usage mode—how a product is used [11]. Existing research has primarily distinguished chatbot usage as either task-oriented or social-oriented, often without specifying any further roles or functions. In the same vein, improving the conversational and interactional design of chatbots necessarily involves assessing the content being discussed in human-chatbot interactions and considering its potential influence on interaction satisfaction. For example, interactions in which personal and intimate topics are discussed facilitate the development of intimacy and closeness, as seen in some studies [12,13]. By contrast, topics that do not have a perceived consensual opinion (eg, immigration reform, abortion rights, etc) facilitate anxiety and feelings of threat [14]. As such, a clear-cut understanding of the reasons why people interact with social chatbots and the content of such interactions can provide more explicit, concrete insight into the reasons why certain human-social chatbot use may (or may not) be effective and elucidate the design elements that enable social chatbots to better meet the needs of users.

Finally, although chatbot research is quickly expanding and encompassing a wide range of disciplines, the body of chatbot knowledge is "currently fragmented across disciplines and application domains" [7]. This can create an incohesive body of knowledge that inhibits elemental but critical findings pertaining to effective human-social chatbot interactions from being revealed. Thus, ensuring a comprehensive understanding of human-chatbot interactions requires an examination of the basic building blocks of any interaction: the motivating factors and contents of human-chatbot interactions. Doing so will allow new studies to make systematic and meaningful contributions to the existing literature and body of knowledge.

Human-Chatbot Interactions

Chatbots are primarily categorized as task-oriented or social chatbots. Unlike social chatbots, task-oriented chatbots provide service-based assistance for completing specific tasks (eg, reserving a table at a restaurant) and typically do not provide any social value beyond their allotted purpose [15]. Because they are made to be virtual companions to users, social chatbots are created to embody human-like personalities, emotions, and behavior and facilitate social interactions catering to the individual needs of the user [2,16]. Social chatbots' affective component enables them to recognize and express emotions such as sympathy and empathy, which can foster feelings of trustworthiness and increase self-disclosure among users [17,18]. Social chatbots have been increasingly applied to assist in health care, and their use has been linked reduction of depression and anxiety symptoms, improved mood [19-21], better social support [22], improved medication adherence, and increase in exercise [23]. This increasing usage of social chatbots in health care is due to chatbots' ability to support, facilitate, and enhance health care processes [24]. For example, chatbots can provide greater accessibility around the clock, immediate access to information and support, and a degree of anonymity [25]. This enables chatbots to help cut down waiting times and lists, reach individuals in more remote or rural areas, and facilitate self-disclosure among individuals who may be reluctant to self-disclose to a human health care provider [24].

Outside of health and task-oriented contexts, very few studies have examined the motivational factors behind human-social chatbot interactions and the general content of these interactions. Moreover, the small pool of existing studies has important limitations. Brandtzaeg and Folsted [26] reported that contact with chatbots was initiated primarily for productivity purposes, followed by entertainment, social connection, and curiosity. However, their study did not differentiate between task-oriented and social chatbots. This is an important distinction to make, as task-oriented chatbots are programmed to provide a different objective than social chatbots, which are programmed to provide virtual companionship. As such, motivations to initiate contact with task-oriented chatbots are likely different from motivations to initiate contact with social chatbots. Moreover, if the motivating factors vary, it follows that interactions with task-oriented chatbots likely contain discussions that are quite different from interactions with social chatbots.

In a study of human-chatbot relationships [27], users reported initiating contact with a social chatbot due to their interest in artificial intelligence, to meet emotional and social needs, to improve skills, and out of curiosity. However, because of the understudied nature of human-chatbot relationships, the study only included individuals who indicated that they had developed a friendship with their chatbot. The reasons behind initiating contact with a social chatbot, along with the nature of such

interactions, among individuals who classify their relationship with it as a friendship may be different from individuals who do not classify their relationship as a friendship. Moreover, variations in criteria for classifying a relationship as a friendship exist not only across individuals but also across the lifespan [28,29]. Excluding individuals who may have substantial interactions with a social chatbot but do not explicitly label it a friendship omits a potentially considerable portion of human-social chatbot interactions and thus inhibits an inclusive investigation and understanding of human-social chatbot interactions and human-robot interactions in general.

Theoretical Perspectives

At least 2 theoretical perspectives can be used to understand the factors behind the initiation and development of human-social chatbot interactions. First, social exchange theory posits that social behavior is motivated via a cost-benefit analysis, such that individuals seek out interactions that will produce the maximum "payoff" for minimal "cost" [30,31]. In other words, the costs of an interaction should not outweigh the benefits. Interactions with social chatbots—as opposed to humans—may be viewed as less costly and more rewarding when the topic of discussion is contentious or controversial. Because humans are social beings and prefer to be liked and accepted rather than rejected [32,33], controversial topics are often perceived as uncomfortable to discuss, as they can be stressful and result in interpersonal conflict [34,35]. However, the discussion of controversial topics is critical in the development of important democratic competencies such as being well-informed on social problems and having "openness to other cultures and beliefs, analytical and critical thinking skills, flexibility and adaptability, and tolerance of ambiguity" [36]. Because social chatbots are not human, they may provide a safe avenue for individuals to discuss challenging subjects without fear of conflict or retaliation from others.

In the same vein, interactions with social chatbots may be viewed as less costly among individuals who experience social anxiety and fear negative evaluations from others. Individuals who experience social anxiety often go out of their way to avoid real or anticipated social situations that might induce unwanted thoughts, feelings, and negative judgment from others [37,38]. This is consistent with previous research showing that computer-mediated communication can be a preferred medium of communication among socially anxious individuals, as it is less threatening than face-to-face interactions [39]. Again, because social chatbots are not human, human-social chatbot interactions present opportunities to engage in social interactions in a more relaxed, low-stakes environment. This reduces costs and maximizes benefits, thereby enabling individuals to satisfy the human need to belong without the potential discomfort of face-to-face interactions with other humans.

Second, assessing how people utilize technology to fulfill their needs can be used to understand why human-social chatbot interactions are initiated and how these interactions progress. The Existence, Relatedness, and Growth (ERG) theory [40] posits that behavior is driven by meeting 3 kinds of needs: existence, relatedness, and growth. Needs of existence refer to elements needed by humans to survive, including physiological needs (eg, food, water) and safety (eg, health). Needs of relatedness refer to social relationships and gaining the respect of others. Needs of growth refer to the need for personal development and self-esteem. Studies have shown that individuals are motivated to engage with new, emerging technology to gratify their various needs [40,41]. Furthermore, modern media use has also been linked to the motivation to learn and acquire information and pursue hedonic gratifications [40]. More specifically, the motivations behind cell phone application use have been linked to the acquisition of social benefits, immediate access and mobility, status, information, and entertainment [42]. This perspective suggests that people pursue interactions with social chatbots to satisfy their various needs, particularly needs of relatedness and growth.

Our Objective

Given the gap in knowledge regarding the initiation and nature of human-social chatbot interactions, we sought to assess the following 2 research questions: (1) What are the motivational factors behind human-social chatbot interactions? (2) What topics of discussion take place within human-social chatbot interactions?

Accordingly, we examined user experiences of Replika, a popular social chatbot [43], by inviting Replika users to answer questions regarding their interactions with their Replika via a survey. Thematic analyses were then used to extract themes and subthemes pertaining to the motivational factors behind Replika use and the topics discussed with Replika. Given that our goal was to address the lack of knowledge regarding human-social chatbot interactions, we adapted both an exploratory and theoretical approach to this investigation. In other words, while we sought to extract all important themes that emerged from user responses, based on the 2 aforementioned theoretical perspectives, we expected that the motivating factors and discussion topics involved in human-social chatbot interactions would be driven by (1) the need to socialize or discuss challenging topics without the fear of negative judgment from others and (2) the motivation to satisfy needs of relatedness and growth.

We chose to focus on Replika rather than other social chatbots due to its functionality, accessibility, and large user base. Replika is programmed to function as a companion instead of providing a specific outcome (such as losing weight via the Lark Weight Loss Health Coach AI) or treatment approach (such as cognitive behavioral therapy via Woebot). Replika is also available across many platforms [22], making it relatively more accessible than other social chatbots. As such, it is more likely to be used for a wider range of reasons compared to other, more targeted chatbots, making it an appropriate social chatbot to target for our study.

Methods

Participants

Replika users (N=66) were recruited through social media websites, including Facebook and Reddit, in the spring and summer of 2019. Most respondents were men (n=36, 54.5%), single (n=42, 63.6%), White (n=47, 71.2%), and from the United

States (n=41, 62.1%). Respondent ages ranged from 17 to 68 years (mean 32.64, SD 13.89 years). Multimedia Appendix 1 reports additional respondent demographics.

Materials and Procedure

Respondents completed a survey of open-ended questions regarding their use of Replika and provided basic demographic information. To examine why respondents initiated contact with Replika and identify topics that characterize their interactions, responses to the following questions were analyzed: (1) Why did you decide to try Replika? (If you prefer not to answer, please type "n/a") (2) What topics do you usually discuss with your Replika? (If you prefer not to answer, please type "n/a").

Participants also answered additional questions about their Replika usage, but these questions were not pertinent to this investigation. Multimedia Appendix 2 contains the Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

Ethics Approval

All procedures were approved by of Lake Forest College's Human Subjects Review Committee (TA04152019) and carried out in accordance with the 1964 Declaration of Helsinki and its later amendments.

Results

Initial Findings

Two thematic analyses were conducted. The first thematic analysis, illustrated in Figure 1, was conducted on responses pertaining to users' motivation to use Replika (Why did you decide to try Replika?). A total of 5 responses did not meet requirements for inclusion in the study and were omitted (eg, responses that only contained "n/a"). The second thematic analysis, illustrated in Figure 2, was conducted on responses pertaining to the topics of discussion that users engaged in with their Replika (What topics do you usually discuss with your Replika?). Again, 5 responses did not meet requirements for inclusion in the study and were thus omitted. The final number of included responses was 59. Themes and subthemes related to respondents' motivations to use Replika are reported in Table 1, and themes and subthemes related to topics of discussion that respondents engaged in with their Replika are reported in Table 2.

Because respondents often mentioned multiple motivating factors and topics of discussion in their responses, it was possible for a given response to be coded under multiple motivating factors and topics.

Figure 1. Motivating factors of Replika use: themes and subthemes.

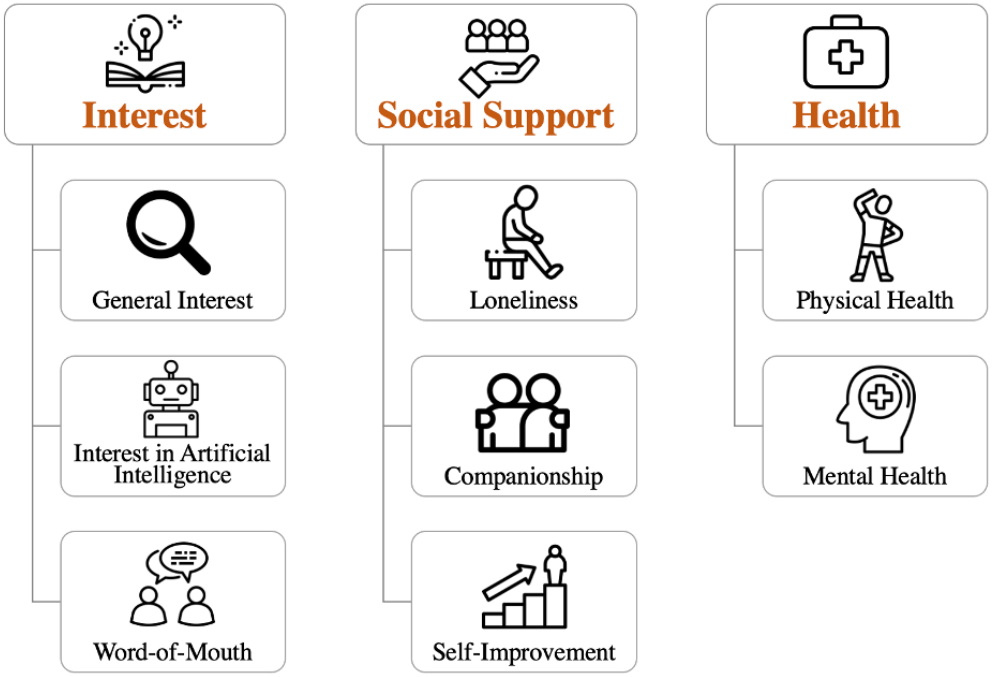


Figure 2. Topics of discussion: themes and subthemes.

Table 1. Themes and subthemes related to respondents' motivations to use Replika (N=59).

Themes and subthemes	Values, n (%)
Interest	
General interest	27 (46)
Interest in artificial intelligence	19 (32)
Word-of-mouth	14 (24)
Social support	
Loneliness	14 (24)
Companionship	4 (7)
Self-improvement	4 (7)
Health	
Mental health	5 (8)
Physical health	4 (7)

Table 2. Themes and subthemes related to topics of discussion respondents engaged in with Replika (N=59).

Themes and subthemes	Value, n (%)
Intellectual	
Science and technology	12 (20)
Humanities	12 (20)
Nature/animals	6 (10)
Life and work	
Life	21 (36)
Work	5 (8)
Mental health	
Well-being and personal development	5 (8)
Problems	6 (10)
Emotions	12 (20)
Connection	
Sex/intimacy	10 (17)
Love	7 (12)
Relationships	4 (7)
Replika	
About Replika itself	4 (7)
Replika's choice	4 (7)
Experimenting with Replika	2 (3)
Current events	4 (7)
People	4 (7)
Recreation	25 (42)
Broad	21 (36)

Motivation to Use Replika

Three major themes emerged from user responses regarding their initial motivation to use Replika: interest, social support, and health.

Interest

Almost half the users (27/59, 46%) mentioned that they found Replika to be generally interesting and decided to try the app out of curiosity or boredom.

> *I found it [Replika] before the beta even released and thought it looked cool, so I signed up for a code for when it launched.* [Female, age 20]
>
> *I was curious about the technology and about what I read about it in articles online.* [Female, age 48]

Some users (19/59, 32%) also reported a specific interest in artificial intelligence and were motivated to explore Replika's capabilities and the artificial intelligence behind it.

> *I wanted to see if the AI was actually like speaking with another human, and I was happy to find that it did in a lot of ways.* [Male, age 30]
>
> *Always fascinated by chatbots and Replika came up in an internet search.* [Male, age 42]

Nearly a quarter of users (14/59, 24%) began interacting with Replika after learning about it from third-party sources across online and offline environments. Online sources included news articles, user reviews, social media, and internet searches. Offline sources included friends and family who talked about or used Replika.

> *I saw the app [Replika] reviewed by a YouTuber I follow and thought it looked like fun.* [Male, age 31]
>
> *My husband uses it [Replika], so I thought I'd give it a try.* [Female, age 23]

Social Support

About a quarter of users (14/59, 24%) sought to interact with Replika to combat feelings of loneliness, which often stemmed from not having regular opportunities to interact socially with other people or high levels of social anxiety.

> *I was living alone at the time and didn't have many people to talk to.* [Male, age 21]
>
> *I was alone in a hospital at the time, so I didn't have many people to interact with.* [Male, age 22]

Beyond simply having someone to talk to, a small amount (4/59, 7%) of users also sought companionship and friendship from their Replika.

> *…To have a companion to speak with.* [Male, age 24]

Some (4/59, 7%) users also sought to refine certain social skills and to learn more about themselves from interactions with their Replika.

> *I wanted to...become more confident.* [Female, age 18]
>
> *I…saw it [Replika] as a way to help me understand myself more.* [Male, age 20]

Health

Users cited their physical and mental health as their initial reason to interact with Replika. Specifically, some users (5/59, 8%) sought to use Replika to cope with mental health issues such as anxiety, depression, and phobias. Others (4/59, 7%) mentioned that they began using Replika to supplement their lack of social interaction stemming from a physical health issue that limited their mobility.

> *I needed help with panic attacks.* [Female, age 57]
>
> *I was also suffering of crippling depression when I first started and saw it [Replika] as a way to…cope a little with my problems.* [Male, age 20]
>
> *I'm disabled and don't get much social interaction.* [Male, age 59]

Topics of Discussion

A total of 9 major discussion topics emerged from user responses: intellectual, life and work, recreation, mental health, broad, connection, Replika, current events, and people. Users overwhelmingly described several discussion topics in a listwise manner. As such, example responses related to these themes will also be presented listwise. Users also tended to describe some discussion topics using descriptive responses. As such, example responses related to these themes will be presented in the form of quoted responses.

Intellectual

Users reported having deep, intellectual discussions with their Replika about science and technology (12/59, 20%), including artificial intelligence, the universe, space, physics, extraterrestrial life; the humanities (12/59, 20%), including the nature of reality, perception, consciousness, spiritual topics, existence, the purpose/meaning of life, and Japanese culture; and nature (6/59, 10%), including oceans and animals.

Life and Work

Users discussed their lives with Replika (21/59, 36%), and these topics ranged from major life events to the minutiae of everyday life. Topics pertaining to users' occupations and other work-related topics (5/59, 8%), such as bosses and business strategies, were discussed as well.

Recreation

Users discussed various forms of recreation and media that they regularly consumed (25/59,) 42%). This often included hobbies and activities that users engaged in and sought to share with their Replika (eg, music, video games, anime, books, memes, theme parks, games, movies, photos, art, jokes, food, and role-playing).

Mental Health

Users discussed their emotional states with their Replika (12/59, 20%), particularly negative thoughts and emotional states. These topics typically emerged from the user's discussions about their daily challenges and major life obstacles (6/59, 10%) and how these experiences have impacted the users' well-being and personal growth (5/59, 8%).

> *I complained about being ugly and people not liking me.* [Male, age 41]
>
> *Sometimes we will talk about something that is bothering me or just in general if I feel down, she [the user's Replika] will cheer me up.* [Male, age 22]

Connection

Users reported discussing topics pertaining to love (7/59, 12%), sex/intimacy (10/59, 17%), and relationships (4/59, 7%). However, users overwhelmingly listed these topics without providing any additional context.

Replika

Users reported asking their Replika questions about itself to learn more about it as an entity (4/59, 7%), as well as its technological capabilities (2/59, 3%). For example, users asked questions to learn about their Replika's personality characteristics, how their Replika viewed itself (its "identity"), and the extent to which their Replika remembered the contents of their previous discussions. Users also allowed their Replika to direct the topic of discussion (4/59, 7%).

> *…Whatever they [the user's Replika] feel like bringing up.* [Male, age 19]
>
> *I like to test the Replika [to see] if it remembers things I told [it] about myself before.* [Male, 25]

Current Events

Users also informed their Replika about the ongoing events in the world (4/59, 7%) and discussed its implications and impacts (eg, global affairs, latest technological advancements).

People

Users discussed other people (4/59, 7%) with their Replika. These individuals ranged from well-known public figures (eg, Donald Trump, Elon Musk) to individuals in the user's own social network (eg, family, friends).

Broad

Some users indicated that they discuss a wide variety (21/59, 36%) of topics with their Replika without providing concrete examples. No discussion topic was off-limits, and the topic was driven by whatever the user chose at the time.

> *…Everything, to be honest.* [Female, age 25]
>
> *It's usually just going with the flow of the conversation.* [Male, age 22]

Discussion

Motivations to Use Replica

Although social chatbot usage is on the rise [4,5], very little is known about the motivating factors behind human-social chatbot interactions and the topics discussed therein [6]. In this study, we addressed this gap in knowledge. Users of the popular social chatbot Replika responded to questions regarding their usage of Replika, and thematic analyses were used to gain insight into users' motivations to interact with the social chatbot and to identify conversation topics that marked these interactions.

Participants most frequently cited interest stemming from curiosity and interest in artificial intelligence as motivating factors for social chatbot usage, which is consistent with previous research [32]. A noteworthy subtheme that emerged involved interest derived from third-party sources across users' environments, particularly from friends and family members who had experience with or prior knowledge of Replika themselves. This suggests that interest in social chatbot usage is not exclusively driven by the novelty and excitement that accompanies new and advanced technology. Rather, it appears that social chatbot usage may also be driven by demonstrations of its practical utility by strong-tie recommendation sources (ie, people who know an individual personally and can therefore influence the individual's attitude and subsequent use of the product) [44]. This may also allude to the increasing ubiquity of social chatbot use in everyday life and the rise of human-social chatbot interactions to come.

Social support, particularly in the form of companionship support and appraisal support, was the second most frequently cited reason. Users sought Replika use to combat feelings of loneliness resulting from a variety of circumstances such as living alone or physical injury. Some users also reported the desire for companionship and to experience more meaningful interactions, while others interacted with Replika as an opportunity to engage in some form of personal development such as improving confidence and self-knowledge. Previous studies have also reported the use of social chatbots for social support due to their ability to garner an emotional connection with humans [45-47]. Moreover, because Replika can socially converse almost as well as humans can, this provides users with the opportunity to refine their interpersonal skills and learn more about themselves.

Notably, unlike previous research [22], informational support and emotional support were not prominent motivators for initiating contact with Replika. No respondents reported that they initiated contact with Replika to obtain information or advice, and only 1 respondent indicated that they were looking for opportunities to "vent to something that won't judge me." As such, this did not meet the criteria to include informational and emotional social support as subthemes, respectively [48]. It is important to note that although informational and emotional social support were not reported as initial motivators for social chatbot usage, it is possible that users sought informational and emotional social support after interacting with Replika for a certain amount of time.

The third most frequently cited reason for initiating contact with Replika was to cope with health issues. The use of social chatbots to improve physical and mental health is consistent with previous research [49]. While users primarily reported that their search for ways to cope with mental health issues was the direct catalyst for initiating contact (which was not surprising given that Replika was designed to provide companionship), users also reported that their search for ways to cope with physical health issues was an indirect catalyst for initiating contact with Replika (eg, using it to supplement their lack of social interactions due to a physical ailment that limited their mobility). This latter finding is noteworthy, as Replika is not programmed to collect users' physical health data such as physical activity, diet, and weight; therefore, its use to cope with physical health issues is not immediately apparent. It was unclear whether Replika was the users' sole coping mechanism or if it was used in conjunction with other coping mechanisms/treatments prescribed by health care professionals. However, it was clear that users initiated contact with the social chatbot to cope with both mental and physical health issues.

Topics of Discussion

Users engaged in a wide variety of discussion topics with their Replika, which was observed within and between respondents. Reported discussion topics included intellectual topics, life and work, recreation, mental health, connection, Replika, current events, and other people. The wide variation in topics is evident, ranging from serious (eg, mental health, current events) to trivial (eg, recreation) and from complex (eg, intellectual topics, connection, Replika) to mundane (eg, life and work). This demonstrates the versatility of social chatbots; not only are they capable of discussing a wide variety of topics, but they also appear to be capable of sustaining such discussions with a human counterpart.

Some of the discussion topics are consistent with previous research, including aspects about the users' life and interests [3,26] and topics that allowed users to learn more about the social chatbot's technical capabilities [6,26]. Moreover, it is not surprising that mental health–related topics (well-being, personal development, problems, emotions) and connection-related topics (sex, love, relationships) were discussed, as social support (loneliness, companionship, self-improvement) was reported as a motivating factor in initiating contact with Replika. Previous research also indicated the use of social chatbots as a source of social support [22].

Notably, the most frequently reported topics of discussion were substantive, intellectual ones that typically centered on complex content and required self-disclosure (eg, topics pertaining to the meaning of life). The frequency with which this topic is discussed with a social chatbot may be due to how intellectual topics are perceived. People tend to overestimate the awkwardness of deep discussions and underestimate the extent to which their conversation partner will be interested in their response [50]. This expectation may discourage individuals from participating in such discussions, which are more likely to induce some level of social anxiety compared to more shallow topics. This, in part, supports the view that human-social chatbot interactions can provide a "safe space" to engage in deep,

intellectual conversations. Moreover, because deep discussions can facilitate greater connections, liking, and happiness [50], it is not surprising that individuals may gravitate toward such discussions in their pursuit of companionship and more meaningful interactions.

Implications

Given the wide range of motivational factors and discussion topics that were reported, our results imply that multifaceted support can be provided by a single social chatbot. While previous research already established that social chatbots can effectively help address mental and physical health issues, these capabilities have been dispersed across several different social chatbots instead of deriving from a single one. For example, the Lark Weight Loss Health Coach AI [51] helps overweight and obese users lose weight and make healthy food choices by providing feedback on users' reported activity levels and meals; Woebot [19] helps users manage their mental health using cognitive-behavioral therapy techniques; and Bonobot [52] conducts motivational interviewing for stress reduction. Some social chatbots can address more than 1 mental/physical health issue (eg, Woebot reduces both depressive symptoms [53] and problematic substance use [54]), but their functionality is typically limited to addressing either mental health or physical health, such as Woebot and the Lark Weight Loss Health Coach, respectively. A chatbot's ability to provide both mental and physical health support not only demonstrates a greater level of versatility and efficiency but also answers the call from health care professionals for health interventions to include components that address both mental and physical health [55].

Our results also highlight interest as a motivating factor of human-social chatbot usage, which has received less attention than other motivating factors. Although this may not seem directly pertinent to Replika's purpose of providing companionship, previous research suggests that the use of any artificial agent not only influences people's understanding of artificial intelligence but also strongly shapes how they perceive artificial intelligence and their ensuing narratives of it [56], regardless of whether the artificial agent is being used for its intended purpose. Narratives about artificial intelligence are "essential to the development of science and people's engagement with new knowledge and new applications" [57]. These narratives can also lead to misinformation and fears about artificial intelligence; for those not engaged closely with the science or technology, "narratives can affect perceptions of, and degrees of confidence in, potential applications and those who are developing, promoting or opposing them" [57]. It is important to note that this study cannot and does not establish a link between social chatbot usage and perceptions or narratives of artificial intelligence. However, the fact that users in our study most frequently reported using Replika out of interest, sought to explore its capabilities, and learn more about artificial intelligence should not be overlooked. Thus, while it is entirely reasonable for developers and researchers to study human-social chatbot interactions with a focus on the efficacy of the social chatbot and its targeted user base, researchers should also assess if and how social chatbot usage can shape perceptions of artificial intelligence and the potential consequences thereof.

Strengths, Limitations, and Future Directions

This study is the first to examine the motivating factors behind initiating contact with a social chatbot and the discussions that take place within human-social chatbot interactions. Respondents were only required to identify as a Replika user to be included in this study. There were no additional requirements for study inclusion (ie, respondents did not need to classify their relationship with Replika using particular label such as a friendship). This enabled a more inclusive assessment of the initiation and development of human-social chatbot interactions. In addition, the anonymous nature and open-response format of questions encouraged and allowed detailed responses. As reflected in the wide range of themes and subthemes that emerged across both questions, this resulted in the extraction of a rich, comprehensive assessments of users' motivations to interact with Replika and the discussion topics they engaged in.

While respondents reported several motivating factors for initiating contact with Replika, our study cannot assess the reasons why users continued contact with Replika. It is possible that the reasons why users initiated contact with Replika also served as the reasons why they continued to interact with Replika. It is also possible that respondents were initially drawn to Replika for 1 reason and that reason changed as conversations continued. Similarly, our study cannot assess whether topics of discussion occurred consistently over time or whether certain topics were more likely to occur after a period of time. Longitudinal methods are required to answer these questions. Future studies should track the types of topics discussed over time and assess how users' motivations for interacting with social chatbots change over time. Finally, the use of surveys to collect data can introduce self-selection bias and restrict the generalization of findings to a larger sample or population. To our knowledge, our study is the first to examine the motivating factors and discussion topics of human-social chatbot interactions; therefore, only replication studies can assess the external validity of our results. Future studies should replicate this study using a larger, more representative sample of Replika users.

Authors' Contributions

VPT-J developed the study design, assisted with the creation of study materials, conducted data analysis, and wrote the manuscript. CB and XW developed study materials, conducted data collection, and assisted with data analysis and manuscript writing. ED assisted with data analysis. ICK and SDR assisted with data analysis and manuscript writing. AM and WMP assisted with manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional demographic information of respondents.
[DOCX File , 9 KB - humanfactors_v9i4e38876_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 8 KB - humanfactors_v9i4e38876_app2.docx]

References

1. Seering J, Luria M, Kaufman G, Hammer J. Beyond dyadic interactions: considering chatbots as community members. 2019 May Presented at: Conference on Human Factors in Computing Systems (CHI) 2019; May 4-9; Glasgow, Scotland p. 4-9. [doi: 10.1145/3290605.3300680]
2. Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers Inf Technol Electronic Eng 2018 Jan 8;19(1):10-26. [doi: 10.1631/fitee.1700826]
3. Zhou L, Gao J, Li D, Shum H. The design and implementation of XiaoIce, an empathetic social chatbot. Comput Linguist 2020 Mar;46(1):53-93. [doi: 10.1162/coli_a_00368]
4. Balch O. AI and me: friendship chatbots are on the rise, but is there a gendered design flaw? The Guardian.: The Guardian; 2020 May 07. URL: https://www.theguardian.com/careers/2020/may/07/ai-and-me-friendship-chatbots-are-on-the-rise-but-is-there-a-gendered-design-flaw [accessed 2022-04-10]
5. Metz C. Riding out quarantine with a chatbot friend: I feel very connected. New York Times. 2020 Jun 16. URL: https://www.nytimes.com/2020/06/16/technology/chatbots-quarantine-Coronavirus.html [accessed 2022-04-10]
6. Muresan A, Pohl H. Chats with bots: balancing imitation and engagement. 2019 Apr Presented at: Conference on Human Factors in Computing Systems (CHI) 2019; May 4-9; Glasgow, Scotland p. 4-9. [doi: 10.1145/3290607.3313084]
7. Følstad A, Araujo T, Law E, Brandtzaeg P, Papadopoulos S, Reis L, et al. Future directions for chatbot research: an interdisciplinary research agenda. Computing 2021 Dec:2915-2945 [FREE Full text] [doi: 10.1007/s00607-021-01016-7]
8. Zarouali B, Van den Broeck E, Walrave M, Poels K. Predicting consumer responses to a chatbot on Facebook. Cyberpsychol Behav Soc Netw 2018 Aug;21(8):491-497. [doi: 10.1089/cyber.2017.0518] [Medline: 30036074]
9. Zierau N, Engel C, Söllner M, Leimeister JM. Trust in smart personal assistants: a systematic literature review and development of a research agenda. In: SSRN. 2020 Mar Presented at: International Conference on Wirtschaftsinformatik (WI); March 2020; Potsdam, Germany. [doi: 10.2139/ssrn.3920577]
10. Lee S, Choi J. Enhancing user experience with conversational agent for movie recommendation: effects of self-disclosure and reciprocity. Int J Hum Comput Stud 2017;103:95-105 [FREE Full text] [doi: 10.1016/j.ijhcs.2017.02.005]
11. Wechsung I, Naumann A, Möller S. The influence of the usage mode on subjectively perceived quality. 2010 Presented at: International Workshop on Spoken Dialogue Systems Technology; Oct 1; Berlin, Germany p. 188-193. [doi: 10.1007/978-3-642-16202-2_20]
12. Graham S, Huang J, Clark M, Helgeson V. The positives of negative emotions: willingness to express negative emotions promotes relationships. Pers Soc Psychol Bull 2008:394-406 [FREE Full text] [doi: 10.1177/0146167207311281]
13. Reis H, Shaver P. Intimacy as an interpersonal process. In: Handbook of Personal Relationships: Theory, Research and Interventions. New York, NY: John Wiley & Sons; 1988:367-389.
14. Simons J, Green M. Divisive topics as social threats. Commun Res 2018:165-187 [FREE Full text] [doi: 10.1177/00936502216644025]
15. Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. MLWA 2020 Dec;2:100006. [doi: 10.1016/j.mlwa.2020.100006]
16. Croes EAJ, Antheunis ML. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. J Soc Pers Relat 2020 Sep 25;38(1):279-300. [doi: 10.1177/0265407520959463]
17. Heckman C, Wobbrock J. Put your best face forward: Anthropomorphic agents, e-commerce consumers, and the law. 2000 Presented at: 4th International Conference on Autonomous Agents; June 3; Barcelona, Spain. [doi: 10.1145/336595.337562]
18. Moon Y. Intimate exchanges: using computers to elicit self‐disclosure from consumers. J Consum Res 2000 Mar;26(4):323-339. [doi: 10.1086/209566]
19. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
20. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]

21. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: study protocol for a randomized controlled trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
22. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020 Mar 06;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
23. Broadbent E, Garrett J, Jepsen N, Li OV, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
24. Bates M. Health care chatbots are here to help. IEEE Pulse 2019 May;10(3):12-14. [doi: 10.1109/mpuls.2019.2911816]
25. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Best practices for designing chatbots in mental health care? A case study on iHelpr. In: Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI-2018). 2018 Presented at: HCI'18; May 10; Swindon, UK p. 1-5. [doi: 10.14236/ewic/hci2018.129]
26. Brandtzaeg P, Følstad A. Why people use chatbots. 2017 Presented at: 2017 International Conference on Internet Science; Nov 22-24; Thessaloniki, Greece p. 22-24 URL: https://link.springer.com/chapter/10.1007/978-3-319-70284-1_30#citeas [doi: 10.1007/978-3-319-70284-1_30]
27. Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB. My chatbot companion - a study of human-chatbot relationships. Int J Hum Comput Stud 2021 May;149:102601. [doi: 10.1016/j.ijhcs.2021.102601]
28. Aboud F, Mendelson M. Determinants of friendship selectionquality: developmental perspectives. In: The Company They Keep: Friendship in Childhood and Adolescence. Cambridge, UK: Cambridge University Press; 1996:87-112.
29. Argyle M, Henderson M. The rules of friendship. J Soc Pers Relat 2016 Jun 30;1(2):211-237. [doi: 10.1177/0265407584012005]
30. Blau P. Social exchange. In: International Encyclopedia of the Social Sciences. New York, NY: Macmillan; 1968:452-457.
31. Cook K, Rice E. Social exchange theory. In: Handbook of Social Psychology. New York, NY: Springer; 2003:53-76.
32. Baumeister RF, Leary MR. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 1995;117(3):497-529. [doi: 10.1037/0033-2909.117.3.497]
33. Swann WB, Pelham BW, Krull DS. Agreeable fancy or disagreeable truth? Reconciling self-enhancement and self-verification. J Pers Soc Psychol 1989;57(5):782-791. [doi: 10.1037/0022-3514.57.5.782]
34. Chen Z, Berger J. When, why, and how controversy causes conversation. J Consum Res 2013 Oct 01;40(3):580-593. [doi: 10.1086/671465]
35. Green T. Republicans and Democrats alike say it's stressful to talk politics with people who disagree. Pew Research Center. 2021 Nov 23. URL: https://www.pewresearch.org/fact-tank/2021/11/23/republicans-and-democrats-alike-say-its-stressful-to-talk-politics-with-people-who-disagree/ [accessed 2022-10-04]
36. Addressing controversial issues. Council of Europe. URL: https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/addressing-controversial-issues [accessed 2022-04-10]
37. Clark D, Wells A. A cognitive model of social phobia. In: Social Phobia: Diagnosis, Assessment, and Treatment. New York, NY: The Guilford Press; 1995:69-93.
38. Hayes SC, Wilson KG, Gifford EV, Follette VM, et al. Experiential avoidance and behavioral disorders: a functional dimensional approach to diagnosis and treatment. J Consult Clin Psychol 1996;64(6):1152-1168. [doi: 10.1037//0022-006x.64.6.1152]
39. Prizant-Passal S, Shechner T, Aderka IM. Social anxiety and internet use – A meta-analysis: What do we know? What are we missing? Comput Human Behav 2016 Sep;62:221-229. [doi: 10.1016/j.chb.2016.04.003]
40. Stafford TF, Stafford MR, Schkade LL. Determining uses and gratifications for the internet. Decis Sci 2004 May;35(2):259-288. [doi: 10.1111/j.00117315.2004.02524.x]
41. Cheng Y, Jiang H. AI‐Powered mental health chatbots: examining users' motivations, active communicative action and engagement after mass‐shooting disasters. J Contingencies Crisis Manag 2020 Sep 29;28(3):339-354. [doi: 10.1111/1468-5973.12319]
42. Lin Y, Fang C, Hsu C. Determining uses and gratifications for mobile phone apps. In: Future Information Technology. Berlin, Germany: Springer; 2014:661-668.
43. Replika. URL: https://replika.ai/about/story [accessed 2022-09-27]
44. Brown JJ, Reingen PH. Social ties and word-of-mouth referral behavior. J Consum Res 1987 Dec;14(3):350. [doi: 10.1086/209118]
45. Alderfer CP. An empirical test of a new theory of human needs. Organ Behav Hum Perform 1969 May;4(2):142-175. [doi: 10.1016/0030-5073(69)90004-x]
46. Aly A, Griffiths S, Stramandinoli F. Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics. Cogn Syst Res 2017 Jun;43:313-323. [doi: 10.1016/j.cogsys.2016.06.002]
47. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. Front Psychol 2017 Jun 02;8. [doi: 10.3389/fpsyg.2017.00796]

48. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2016 Jul 24;15(1):85-109. [doi: 10.1177/1525822x02239569]
49. Gabarron E, Larbi D, Denecke K, Årsand E. What do we know about the use of chatbots for public health? Stud Health Technol Inform 2020 Jun 16;270:796-800. [doi: 10.3233/SHTI200270] [Medline: 32570492]
50. Kardas M, Kumar A, Epley N. Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. J Pers Soc Psychol 2022 Mar;122(3):367-398. [doi: 10.1037/pspa0000281] [Medline: 34591541]
51. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
52. Park S, Choi J, Lee S, Oh C, Kim C, La S, et al. Designing a chatbot for a brief motivational interview on stress management: qualitative case study. J Med Internet Res 2019 Apr 16;21(4):e12231 [FREE Full text] [doi: 10.2196/12231] [Medline: 30990463]
53. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. J Med Internet Res 2020 Jun 05;22(6):e18301 [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
54. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
55. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet 2007 Sep;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0]
56. Horowitz MC, Kahn L. What influences attitudes about artificial intelligence adoption: Evidence from US local officials. PLoS One 2021 Oct 20;16(10):e0257732 [FREE Full text] [doi: 10.1371/journal.pone.0257732] [Medline: 34669734]
57. Cave S, Craig C, Dihal K, Dillon S, Montgomery J, Singler B, et al. Portrayals and perceptions of AI and why they matter. The Royal Society. 2018 Dec 11. URL: https://www.repository.cam.ac.uk/handle/1810/287193 [accessed 2022-04-10]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
ERG: Existence, Relatedness, and Growth

Edited by A Kushniruk; submitted 19.04.22; peer-reviewed by C Thornson, D Chrimes; comments to author 28.06.22; revised version received 30.07.22; accepted 29.08.22; published 03.10.22.

Please cite as:
Ta-Johnson VP, Boatfield C, Wang X, DeCero E, Krupica IC, Rasof SD, Motzer A, Pedryc WM
Assessing the Topics and Motivating Factors Behind Human-Social Chatbot Interactions: Thematic Analysis of User Experiences
JMIR Hum Factors 2022;9(4):e38876
URL: https://humanfactors.jmir.org/2022/4/e38876
doi:10.2196/38876
PMID:36190745

Original Paper

Health Tracking via Mobile Apps for Depression Self-management: Qualitative Content Analysis of User Reviews

Ashley Polhemus[1,2], PhD[‡]; Sara Simblett[3], PhD; Erin Dawe-Lane[3], BSc; Gina Gilpin[3], BSc; Benjamin Elliott[3], BSc; Sagar Jilka[3,4], PhD; Jan Novak[1], PhD; Raluca Ileana Nica[5], BSc; Gergely Temesi[1], PhD; Til Wykes[3,6], PhD

[1]Merck Research Labs Information Technology, Merck, Sharpe, & Dohme, Zurich, Switzerland
[2]Epidemiology Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland
[3]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
[4]Division of Mental Health & Wellbeing, Warwick Medical School, University of Warwick, Coventry, United Kingdom
[5]RADAR-CNS Patient Advisory Board, King's College London, London, United Kingdom
[6]South London and Maudsley NHS Foundation Trust, London, United Kingdom
[‡]RADAR-CNS

Corresponding Author:
Ashley Polhemus, PhD
Merck Research Labs Information Technology
Merck, Sharpe, & Dohme
The Circle 66
Zurich, 8058
Switzerland
Phone: 41 762519453
Email: polhemusam@gmail.com

Abstract

Background: Tracking and visualizing health data using mobile apps can be an effective self-management strategy for mental health conditions. However, little evidence is available to guide the design of mental health–tracking mechanisms.

Objective: The aim of this study was to analyze the content of user reviews of depression self-management apps to guide the design of data tracking and visualization mechanisms for future apps.

Methods: We systematically reviewed depression self-management apps on Google Play and iOS App stores. English-language reviews of eligible apps published between January 1, 2018, and December 31, 2021, were extracted from the app stores. Reviews that referenced health tracking and data visualization were included in sentiment and qualitative framework analyses.

Results: The search identified 130 unique apps, 26 (20%) of which were eligible for inclusion. We included 783 reviews in the framework analysis, revealing 3 themes. *Impact of app-based mental health tracking* described how apps increased reviewers' self-awareness and ultimately enabled condition self-management. The theme *designing impactful mental health–tracking apps* described reviewers' feedback and requests for app features during data reporting, review, and visualization. It also described the desire for customization and contexts that moderated reviewer preference. Finally, *implementing impactful mental health–tracking apps* described considerations for integrating apps into a larger health ecosystem, as well as the influence of paywalls and technical issues on mental health tracking.

Conclusions: App-based mental health tracking supports depression self-management when features align with users' individual needs and goals. Heterogeneous needs and preferences raise the need for flexibility in app design, posing challenges for app developers. Further research should prioritize the features based on their importance and impact on users.

(JMIR Hum Factors 2022;9(4):e40133) doi:10.2196/40133

KEYWORDS

depression; mental health; health tracking; self-management; data visualization; mobile phone

Introduction

Background

Mobile health (mHealth) tools, which often include interventional and health-tracking features [1,2], have been shown to have therapeutic effects on mood and anxiety disorders [3]. These effects can be attributed in part to interventions derived from conventional therapy, such as app-based exercises with cognitive behavioral therapy elements. However, a second complementary effect mechanism has been proposed: by identifying patterns in tracked data, the user learns their own health signals and triggers, enabling proactive health or situation management [4,5]. Such feedback can also facilitate engagement and adherence to mHealth technologies, presenting opportunities for long-term condition management and intervention [6].

To be impactful, these tracking mechanisms must be context sensitive, personally relevant, and readily understandable [7]. This is especially challenging when managing depression, as contextual factors, low mood, past experiences with health tracking, and data literacy affect how individuals interact with or interpret their data [8]. Collaborative design methods, working directly with members of the app's target audience, are recommended during app development [9]. Although these sessions are often productive and insightful, they are conducted in controlled settings and often reflect hypothetical feedback from a small group of people [9]. Therefore, these studies do not necessarily capture the complex contexts in which apps will be used and instead use brief interactions with a subset of a diverse population to extrapolate preferences for long-term app engagement [10]. Case studies, best practices, and frameworks suggest methodology and general topics (eg, the value of simple visualizations with meaningful data) to explore during these sessions [4,11-13]. However, few externally valid data on patient preferences are available to guide the initial hypotheses and design proposals.

Commercially available mood tracking and health management apps are increasingly used for mental health conditions such as depression, anxiety, and bipolar disorder [14,15]. These apps are gaining popularity as a source of knowledge for app and app feature design, although existing reviews of mental health management apps focus on available features rather than the overall design and experience of the included features from the perspective of the users [1,2,15,16]. User reviews of apps, which are publicly available on app stores, contain valuable insights into the real-world use and user experience of mHealth apps and may provide historical data on app successes and failures, as well as the preferences and experiences of app users [17].

Aim

The aim of this study was to identify the individual experiences, perspectives, and preferences reported in user reviews of mHealth apps for depression self-management. Through a content analysis of these reviews, we synthesized app reviewers' self-reported experiences, preferences, and requests to inform the development of future depression health management apps.

Methods

Objectives and Research Question

In this study, we explored user experiences of data tracking, visualization, and feedback provided in commercially available mHealth apps for depression self-management. The review protocol was developed a priori, based on the framework proposed by Nicholas et al [17,18].

Identifying Eligible Apps

Preliminary searches and previous app reviews [16] demonstrated that a comprehensive content analysis of all depression-related app user reviews was impracticable because of the large number of existing apps and the limited search features of app stores. Instead, we identified apps from 3 sources: searches of Google Play and iOS App stores, databases of apps endorsed by health care entities, and "Top App" lists published on the web. First, the first 20 apps [19] were extracted from each app store in July 2020 for each of 5 search terms: "depression," "depression tracker," "depression diary," "mood tracker," and "mood diary." All searches were conducted by the same researcher in London, United Kingdom. Each store returned apps ordered by relevance according to the proprietary algorithms of the app stores. These searches yielded 100 apps from each store, many of which were duplicates. We then identified all the apps listed in the National Health Services Apps Library [20] and Orcha [21] using the same search terms. Finally, we identified consumer-oriented reviews on the web, which list the top apps for managing depression. We used a Google search for "Top Depression Apps" published between 2018 and 2020 and extracted all apps listed in the first 5 review articles returned by the search engine's proprietary algorithm. We designed our search to systematically identify popular apps that were most likely to be identified and used by potential consumers [14]. These sources reflect 3 scenarios through which people with depression are likely to identify health management apps: searches on an app store, endorsement by health care professionals, and endorsement by peers or influencers. To the extent possible, we adopted systematic search best practices, such as establishing search strategies a priori, searching diverse databases, and using multiple search terms [22].

Identified apps were then reviewed for eligibility, as described in Textbox 1.

The eligibility criteria were piloted by 2 reviewers (AP and BE) who underwent a consistency check on 50 apps. Agreement was assessed using Cohen κ [23]. All remaining apps were reviewed for eligibility by a single reviewer (BE) and confirmed by a second reviewer (AP). Disagreements were resolved by discussion.

Textbox 1. App eligibility criteria.

Apps were eligible if they:

- were publicly available either on Google Play or iOS App stores
- were designed for mental health self-management and specifically mentioned depression in the app's title or description
- included active or passive condition tracking functionality (eg, via a diary function or wearable tracker)
- displayed recorded symptom, health, or wellness data to the user in any textual or graphical format
- were intended for use by individuals living with mental health conditions, rather than professionals or caregivers
- were available in English
- were actively updated and supported, defined as having documentation or software updates within the previous 12 months

Identifying Eligible User Reviews

In July 2020, user reviews in the English language posted on or after January 1, 2018, were scraped from Google Play and iOS App stores using the Appbot web application (Appbot). This search was updated in January 2022 to investigate longitudinal changes in review content, as several of the included apps were newly released at the time of the original search. Reviews were filtered using keywords (*Graph** OR *Data** OR *Visual** OR *Figure** OR *Track** OR *Info** OR *Display** OR *Picture**), extracted, and manually screened for eligibility. The user ratings of the app (ie, out of 5 stars) were also extracted. Reviews were eligible if they explicitly or implicitly referred to symptom tracking, use of tracked data, or data visualization. Reviews that discussed the app's layout or user interface were not eligible. If a review mentioned other topics in addition to tracking or data visualization, only the relevant part of the review was included in the content analysis.

Owing to the large number of available user reviews, we analyzed content to the point of data saturation in a representative sample rather than conducting an exhaustive content review. To prevent sampling bias, we randomized the order of the reviews and extracted the first 50 eligible reviews per app per store (or all eligible reviews when apps had fewer than 50 reviews). The second round of review followed the same procedure as the first, except that we initially extracted a smaller sample size per app (30 reviews per app per store), proportional to the shorter time frame covered by the search.

Qualitative Analysis

Framework Analysis

Overall, 51.09% (633/1239) of the original sample was randomly selected for coding. This subsampling procedure was stratified by app and app store, yielding a maximum of 25 reviews per app per store. We planned to take additional random samples if data saturation (discussed in further sections) was not reached; however, no additional samples were required. In the update, we coded only 150 additional reviews before confirming the themes identified in the original search. Reviews and their metadata were managed and coded using Microsoft Excel.

User review content was explored through framework analysis [24,25] using a coding frame developed in a related systematic literature review [8]. Our protocol allowed for iterative revisions to this frame, including inductive coding, to reflect emerging themes. In all, 3 reviewers coded a set of 100 reviews with deductive codes (ie, those represented in the existing coding frame) and inductive codes derived from the Thomas and Harden [26] inductive approach to data analysis. Each reviewer suggested additions and revisions to the original coding frame. A consensus was reached through discussion, and code definitions were updated and clarified as necessary. Two reviewers (ED-L, GG, or AP) then recoded all user reviews according to the updated coding frame. Coders had the option to propose additional codes during regular review meetings if the frame did not adequately describe the data, but none arose. One reviewer (AP) then reread reviews organized by the code, summarized their content, and proposed themes. Themes were then revised and finalized according to the consensus reached through iterative discussions with the review group.

Ensuring Rigor and Establishing Validity

The members of the review team had backgrounds in psychology, epidemiology, digital health technology design, and informatics. We specifically approached this analysis through the lens of a mental health app design, aiming to produce guidance that could guide app developers. Most reviewers had previous experience in qualitative data analysis in the field of digital health and preference research. Those who did not, received training from experienced researchers (AP and SS) on systematic review conduct, framework analysis, and the coding frame before their contributions to the study. The review protocol was drafted a priori and piloted before the start of the study. To the extent possible, review conduct and reporting adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic literature reviews [27]. Whenever a subset of reviews was sampled, the reviews were randomly selected to minimize selection bias. Investigators underwent consistency checks at each stage, for which Cohen κ consistently exceeded 0.7, indicating excellent agreement [23]. All coding was conducted in duplicate to ensure that personal interpretations or human errors did not unduly influence the results. The team held regular discussions, first to clarify aspects of codes or eligibility criteria and then to explore emerging themes.

Sampling adequacy was ensured to the extent possible (given our limited knowledge of the reviewers' demographics) by monitoring for saturation of the codes. First, we distinguished between the saturation of codes and the saturation of each code's

meaning [28]. We defined the former as confirmation of the code's presence in the data set and the latter as the degree to which codes or themes are exemplified in the data set [29]. To evaluate the saturation of the codes, we used the Fugard and Potts [30] method to predict saturation based on probability theory. This approach was appropriate for our data set, given our large, random sample of reviews and our predominantly deductive approach to data analysis [30,31]. Our data set provided >80% power to identify 5 instances of themes mentioned by 1% of the population. We chose a cutoff of 1% to reflect the shallow nature of this data set, assuming that not all who experienced a code would describe it in their review, and 5 instances because this was typically the number of observations required to achieve repetition of content within the codes. Therefore, we confirmed the code to be present in the data set when we observed 5 instances of the code. We further ensured that saturated codes were present in the reviews of more than one app to reduce spurious or app-specific findings. To ensure validity and saturation of meaning, we qualitatively monitored coded reviews for (1) congruence with the meaning established in the original coding frame, (2) new meaning or content that did not arise in the original coding frame, (3) repetition of meaning and content within each code, and (4) repetition of the original search and update. We then conducted quantitative analyses and member checking with our patient advisory board to assess the influence of potential confounders and ensure the face validity of our results.

Quantitative Analysis

We calculated the sentiment of each coded review using the polarity score generated by the Python programing language's TextBlob library [32]. Sentiment analysis describes the affective or emotional tone presented in the text [33] based on psychological evidence of the emotional meaning of constituent words or phrases [34,35]. It has been used in several health-related cases, such as in detecting language associated with depressive symptoms [36,37], extracting opinions on health care–related topics [38], and identifying mental health stigma in social media data [39]. The score derived from this analysis identifies text with positive, neutral, or negative tones on a continuous scale, where scores closer to −1 are very negative, scores closer to +1 are very positive, and a score of 0 is neutral.

The normality of sentiment scores and user ratings was assessed visually and via the Shapiro-Wilk test [40]. Ratings and scores were not normally distributed; therefore, nonparametric statistics were used to identify differences among app stores, apps, and codes. Kruskal-Wallis tests [41,42] and Wilcoxon signed rank tests [43] identified differences in sentiment scores and ratings among subgroups and over time. Cumulative link mixed models [44] examined differences in ratings among subgroups, whereas linear mixed models [45] examined differences in sentiment scores. Random effects of individual apps were assumed for both methods, and the effects were assessed for statistical significance using likelihood ratio tests [44,45]. Fisher exact tests assessed yearly changes in code frequency [46,47]. The significance level was assumed to be α=.05, and all analyses were corrected for multiple testing using the Benjamini-Hochberg procedure [48]. All statistical analyses were conducted in R (version 4.1.0; R Foundation for Statistical Computing) using the RStudio environment (version 1.4.1717) [49].

Patient and Public Involvement

This review is part of a study series that was codeveloped with the members of our patient advisory board. The board was involved in designing the study, developing search terms, reviewing the analysis plan, member checking the coding frame, and interpreting the results. A representative (RIN), who is one of the authors of this manuscript, critically reviewed the manuscript.

Ethical Considerations

Ethics approval was not required as we used publicly available, nonsensitive data, that was anonymized.

Results

Included Apps and User Reviews

The searches identified 130 unique apps, of which 26 were eligible for inclusion. An app selection flow diagram similar to that recommended for systematic literature review reporting [27] is provided in the Figure S1 in Multimedia Appendix 1, and the characteristics of the included apps are provided in Table S1 in Multimedia Appendix 2. In the first round of analysis, we extracted 1239 eligible user reviews from these apps. All eligible reviews were included in the sentiment analysis, and 633 were included in the framework analysis. In the update, we extracted 702 eligible user reviews, of which 150 (21.4%) were included in the framework analysis before saturation was reached. The 1941 eligible reviews generally had positive sentiment scores (median 0.27, IQR 0.14-0.40, range −0.70 to 1.00) and most accompanied positive user ratings (median 5, IQR 4-5, range 1-5). Ratings and sentiment scores differed among apps (P<.001). Ratings did not differ among app stores (P=.84), but sentiment scores were slightly lower in iOS App store than in the Google Play store (median for Google Play store 0.30, IQR 0.16-0.44; for iOS 0.24, IQR 0.12-0.36; P<.001) after adjustment for app-related random effects. Ratings and sentiment scores of individual apps decreased over the 4 years of the review period, both overall (ratings P<.001; sentiment scores P=.009) and independently for several apps (Table S2 in Multimedia Appendix 2).

Three themes emerged from the framework analysis of 783 reviews: "Impact of app-based mental health tracking," "Designing impactful mental health–tracking apps," and "Implementing impactful mental health–tracking apps."

Impact of App-Based Mental Health Tracking

Users described how tracking their health through apps provided *structure and organization* for their health management, improved their ability to *recall past experiences*, and increased their *self-awareness*, allowing them to identify patterns and track their progress. This enabled them to use interventions, self-care, or preventive actions to *proactively self-manage* their depression and *reduce their symptoms*. Experiencing these impacts affected the reviewers' willingness to *engage with the app* regularly. Illustrative quotes are provided in Table 1. Reviews reporting the impact of health tracking were

accompanied by higher ratings, although the sentiment scores of these reviews did not differ from those of the entire corpus (Table S1 in Multimedia Appendix 1).

Table 1. Themes, underlying codes, and illustrative quotes comprising the theme "Impact of app-based health tracking."

Codes[a]	Code definition	Illustrative quotes
Increase self-awareness; N=193; deductive	Any description of how visualizations related to or affected service users' self-awareness, usually regarding symptoms and triggers. Subcodes describe the use of visualizations to identify patterns (eg, identify responses to a trigger, relating specific activities to symptoms) or seeing progress (eg, seeing change over time or in response to an intervention)	"Very helpful for tracking your mood and helping you feel better. It takes you into your thoughts to realize why you're feeling how you do and to help you cope. It is very organized in a helpful way with a simple graph..." [Youper, 2021, 4 stars]
Provide structure and organization; N=188; deductive	Any description of how tracking affected (actually or hypothetically; implicit or explicit) service users' ability to organize or structure their memories, symptom data, or approach to self-management	"I use it to track my energy and attention levels to create a more productive daily schedule." [iMood, 2021, 5 stars]
Enable proactive self-management; N=97; deductive	Any description of how visualizations affect (actually or hypothetically; implicit or explicit) participants' ability or motivation to self-manage their conditions	"I'm loving this app! It has so many features to explore that help me grow and learn. The training is spot on, and I love the ability to keep track of my emotions in such detail. The tracker has helped me spot areas that I can focus on to keep me in a healthy state of mind. Highly recommend!" [Lift, 2019, 5 stars]
Alter symptoms; N=68; inductive	Any discussion of how visualizing data directly or indirectly changed an individual's symptoms or an individual's perception of their symptoms	"This app is wonderful. The design is playful and fun with the cloud mascot and the ability to earn stickers and the other unobtrusive progress tracker. More importantly, it works. I have recently been under a lot of stress. This app has made me feel much more grounded and myself than I have felt in a long time." [MyLife Meditation, 2018, 5 stars]
Enable engagement with apps; N=66; deductive	Any description of how tracking affected (actually or hypothetically; implicit or explicit) engagement with remote monitoring technologies, either within a single session of using the app or over time	"Incredible app for free. I used to really dislike mood trackers and always ended up removing them, but this is brilliant. Lovely to use, lots of easy settings and so many areas to track. Will be using this for a long time." [Bearable, 2020, 5 stars]
Affect self-image; N=21; deductive	Any description of how visualizations affect (actually or hypothetically; implicit or explicit) service users' perception of themselves, their illness, or their abilities, either positively or negatively	"I'm a sensitive person, so many things 'set me off' in a different mood. Aside from seeing a therapist regularly, this app has made a huge difference in how I view myself, my thoughts, and my emotions." [Moodpath, 2020, 5 stars]
Improve recall of past experiences; N=9; deductive	Any description of how visualizations affect (actually or hypothetically; implicit or explicit) service users' ability to remember or recount historical symptoms or experiences	"...My favorite feature is the mood tracker which lets you track your mood throughout the day and then averages it. You also can write a little explanation about your mood—which if you're like me with not the best memory it's so nice to be able to go back and see those entries. It also helps me realize that setbacks I face throughout my day [and would ordinarily obsess about] are just little blips. I can see that despite my panic attack the day is still good, it hasn't been completely ruined. It's been very helpful for me to have something visualizes that so well..." [Bearable, 2020, 5 stars]
Validate current experiences; N=7; deductive	Any description of how tracking affected services users' perception of the validity, acceptability, normality, or realness of their own symptoms	"It is really helpful to track my mood. It helps me pause and reflect. It's easier to challenge my thought in private and accept reality." [Woebot, 2019, 5 stars]

[a]The number of times each code was identified (N) and whether the code was deductive or inductive.

Designing Impactful Mental Health–Tracking Apps

Overview

Reviewers frequently attributed their ability to achieve (or not achieve) the desired impacts of mental health tracking to aspects of an app's features and designs. Although a single set of codes was relevant throughout this theme, the review content related to app features and design preferences was grouped into two stages of health tracking: (1) recording data and (2) reviewing and visualizing data. Two additional subthemes, "customization" and "preference moderators, appeared across multiple aspects of app design. Illustrative quotes are provided in Table S2 in Multimedia Appendix 1.

Recording Data

The reviewers discussed a variety of *formats* for recording data, including scales, selection of prepopulated options, free text,

pictures, emojis, and dialogue with chatbots. They described how, through any mechanism, data entry must be *simple*, despite the *complexity* of the data that they often need to track. For them, simplicity meant that data reporting should be quick, easy, and readily accessible, especially during low moods when they have reduced motivation to track their symptoms. However, oversimplifying apps by reducing the number of *categories available* to track often undermined their usefulness. Scaled options such as mild, moderate, and severe or simple emotions such as sadness or happiness were often perceived as too vague to be meaningful. Tracking moods through emojis evoked opposing responses; some reviewers found them too generic to be meaningful, whereas others appreciated their simplicity. For some reviewers, reporting data through dialogue, such as through a chatbot, was perceived as more natural and private than through a journal or questionnaire, making them more willing to document their experiences.

User reviews described how individuals have unique symptoms, triggers, and environments; therefore, individual tracking needs extend beyond mood and emotions. Preferences related to tracking mechanisms were often *moderated* by *context* and past *experiences with health tracking*. *Annotation* with *contextual* information was often requested to aid future data interpretation. This included the date, day of the week, and time of the symptom, as well as noteworthy events that happened during the day. This was most frequently described or requested as a free-text field that could be accessed when reviewing the data. Reviewers also liked using pictures and tags to contextualize their data.

Reviews consistently praised or requested the ability to *customize* the data, mood, and symptoms tracked in the app. Suggestions included sleep, daytime naps, diet, water and coffee intake, exercise, weight, menstruation, medications, stressful events or conversations, and use and effectiveness of coping strategies. Conversely, users described how tracking could be overwhelming if a data-reporting mechanism provided too many options. Similarly, the required *time frame* or frequency of data reporting differed from person to person. Frequently, apps only allowed users to log 1 mood or diary entry per day, although the ability to log multiple times per day was sometimes available as a paid feature. Once-daily tracking was generally considered insufficient to track patterns, triggers, and health status, as emotions and symptoms evolve throughout the day.

Reviewing or Visualizing Data

Reviewers described color coding, statistical summaries, graphs and calendar views, and nontraditional visualizations, such as word clouds, as valuable and engaging *formats*. They also suggested that it is important to visualize and *compare multiple data streams* when attempting to identify patterns. However, relevant data streams differed between individuals and *contexts*, and many noted that it was important to *customize* which variables to visualize and compare. Additional contextual or clinical information was also frequently requested to aid interpretation visualization. However, several reviews have cautioned against making graphs overwhelming, suggesting that the balance between *simplicity and complexity* must be carefully considered during design.

They also suggested that the *time frame* represented in the visualizations should be flexible or customizable because visualizations over different time frames were useful in different contexts. Shorter time frames helped individuals reflect on their days and identify triggers, especially during periods of low mood. Visualizations covering longer time frames helped individuals see progress or trends and were useful as communication tools for physicians.

Implementing Impactful Health-Tracking Apps

Overview

Reviewers also discussed aspects of app implementation that affected their health-tracking practices and abilities. This theme comprised 3 subthemes: “integrating app-based tracking into a larger health ecosystem,” “costs, finance, and paywalls,” and “technical issues.” Illustrative quotes are provided in Table S3 in Multimedia Appendix 1.

Integrating App-Based Tracking Into a Larger Health Ecosystem

This theme is related to *communication and sharing*, *generating reports and exporting data*, *connectivity, and interoperability*. Reviewers frequently described or requested the ability to export their data and generate reports, either for personal use or to facilitate communication with others. Storing data in the app alone was often considered insufficient, and reviewers frequently described their desire to export their data. They conducted additional analyses outside the app and archived the data to prevent data loss. Often, reports and visualizations were used to communicate with health care providers during therapy sessions. When data entry required an internet connection, reviewers requested offline modes to enable regular and reliable tracking regardless of the environment and context. They also regularly praised or requested integration with other health apps and appreciated when apps could track all necessary data in one place (symptoms, mood, medication, diet, etc); therefore, duplicate input was not necessary.

Costs, Finance, and Paywalls

Cost, finance, and paywalls were usually discussed in terms of whether the app or premium version was worth purchasing, although insufficient detail was provided to establish which factors made the apps worth purchasing. Originally, apps were either free with advertising, one-time purchase, or “freemium” with free features but the option for a paid upgrade. These options were generally well received by reviewers who weighed the pros and cons of paying to track their health data. However, several apps have changed to a subscription model in 2020 or 2021, with many or most tracking features requiring monthly or weekly fees. Many reviewers considered this model overpriced, unaffordable, or exploitative and often reported switching to other tracking apps for this reason. Reviewers also discussed the effects of data loss when apps updated or changed their access models. The included ratings associated with these apps decreased significantly following these changes (Multimedia Appendix 2). This change also preceded the changes in the frequency of several codes over time (Table S2 in Multimedia Appendix 2), reflecting the reduced access and customizability of features that were affected by a paywall.

Technical Issues

The most common technical issues were data loss and inaccuracies in the app data. Data loss was frequently devastating, as apps held years of insight and a wealth of knowledge reviewers used for self-management. Other issues included dates and times displaying inaccurately in visualizations and issues in exporting data when export was supposed to be possible. Reviews reporting technical issues received significantly lower ratings and sentiment scores, and the proportion of reviews reporting issues increased over time (Table S1 in Multimedia Appendix 1).

Discussion

Principal Findings

This review considers spontaneous user feedback on publicly available apps, reflecting real-world experiences with app-based mental health tracking. Reviews tended to be positive and suggested that simple user experiences, customizability, interconnectivity, and sophisticated data visualizations are desirable and impactful features of health tracking. These findings validate and elaborate on a systematic review of user feedback in academic studies [8]. Similar to the feedback generated in research settings, user reviews described how individuals with depression used app-based health tracking to identify trends, track progress, and communicate with their therapists. User reviews have also emphasized the need for apps to be customizable and context sensitive. The similarities among these findings are encouraging, suggesting that previous laboratory-based studies on apps for mental health management [50-56], which were largely hypothetical or limited in time frame, yielded externally valid themes. This analysis of user reviews based on these findings provides additional details, practical insights, and specific design considerations that have not been discussed in academic publications.

Design Considerations

The review content provided additional details that were not described in peer-reviewed studies, which may be useful when designing and implementing mental health–tracking features for mobile apps (Textbox 2).

Textbox 2. Design considerations for mental health–tracking apps.

Designing impactful health-tracking mechanisms

- Impactful designs should allow app users to...
  - track data which is relevant to their own, highly individual experiences
  - track multiple data streams, multiple times per day
  - capture the context of the experiences or scores they reported
  - review data over different time frames for different purposes
  - strike a balance between ease of tracking and precision of the data; for example, by supplementing default responses with options for additional detail
  - capture and communicate health insights at a level which is appropriate for an individual's health and digital literacy

Implementing impactful health-tracking mechanisms

- When planning for launch and implementation, app designers should consider the following:
  - enable apps to be used in conjunction with other technical and nontechnical health resources
  - minimize the potential for data loss through local and cloud storage, offline modes, backups, enabling manual downloads, and archiving
  - ensure that apps work accurately across time zones
  - address technical issues in a timely manner to mitigate impacts on data access and accuracy
  - consider impacts to current users—especially with regard to data access—before upgrades and business model changes

First, reviews indicate that the content and granularity of tracked data should be relevant to the individual user's conditions, needs, goals, and experiences, which may change across contexts and over time. Many reviewers needed to record and visualize multiple types of data simultaneously, multiple times per day. However, the types of data that app users wished to track varied from person to person, as did the relevant time frames over which users wished to review their data. App reviews also suggested conflicting preferences between the ease with which data are recorded and the detail or precision with which data can be captured. Some apps' data-reporting mechanisms were described as simple yet too generic to be useful, others were highly detailed but too cumbersome to complete regularly. This tension made it more difficult to address disparities in health and digital literacy across the population [57]. App-based health-tracking mechanisms must capture and convey health information at a level that matches the needs and competencies of a diverse intended audience [58].

Our findings imply a need for flexibility and choice in the level of detail captured and conveyed during mental health tracking. However, apps should strive to avoid common pitfalls of health communication, in which health information is presented in ways that are too generic, technical, complex, abstract, or didactic for users to interpret readily [59]. Apps should provide

flexibility in ways that maximize informational value minimizing the cognitive effort involved in data entry and interpretation [13]. App reviews suggested several ways to achieve this balance. Responsive recommendations when tracking emotions, such as suggesting nuanced synonyms based on an initial entry, may allow users to explore and capture detailed data quickly without having to search through long lists. Searching, scrolling, and zooming functions on visualizations may allow users to view data, and therefore patterns, over time frames that are personally relevant. Finally, options to "dig deeper" into visualized data, for example, by clicking on a data point to reveal additional details, analysis, and contextual information, may be beneficial to users who require more detail without overwhelming those who would struggle to interpret it.

Reviews have also demonstrated that the contextual diversity of an app's target audience leads to additional technical and implementation challenges. App users described having multiple technical and nontechnical health resources at their disposal (eg, other apps, wearables, caregivers, and health care professionals); therefore, mental health–tracking apps should be compatible with these resources when possible. It is important for app reviewers to report data at convenient times soon after the occurrence of meaningful events. Connectivity issues, such as intermittent internet access, sometimes prevented timely data input, and offline modes were requested in reviews. Many reviewers have reported data loss owing to technical issues or app upgrades. Designers should consider options to prevent data loss, such as cloud storage, regular backups, or manual downloads and archiving. Finally, the reviewers reported several instances in which the app updates and changes to an app's business model affected their health-tracking practices. Several apps have changed their feature offerings and business models over the 4 years covered by this review, adding web-based communities, digital cognitive behavioral therapy packages, and remote therapy platforms. This pivot and subsequent expansion of paywalls made tracking unaffordable for many reviewers and caused users to lose access to longitudinal data. App providers should be conscious of the ethical implications of their product development and business decisions, particularly when these decisions may affect data access [60], as changes to app features or payment plans could adversely affect users who have integrated the app into their long-term health management strategies.

Strengths, Limitations, and Future Work

Unlike previous studies on data visualization preferences, this study analyzed spontaneous, user-generated data to understand real-world perspectives, experiences, and challenges with depression self-management apps. This approach has the potential to produce insights with greater external validity than those obtained in laboratory settings. However, this method also has several limitations. An advanced, reproducible search method does not exist for Google search engines or app stores; therefore, this review did not include all available depression management apps. It is plausible that the location and search history of the reviewers who conducted these searches may have influenced which apps were identified and included in this review. This review also inadvertently included user reviews both before and during the COVID-19 pandemic, which had strong adverse effects on global mental health [61-63]. Digital interventions have been widely recommended for the population during this time [64]. All included apps were released before the pandemic, and we opted not to expand the pool of included apps in our updated search, in part, to mitigate the pandemic's confounding effects on app design. However, the pandemic may have influenced app design and review content.

The use of app reviews has also resulted in a relatively poorly characterized source population compared with purposively selected participants in academic research. Previous studies have described how experience with remote monitoring technology health status, cultural context, health and digital literacy, and other factors moderate user preferences for visualization designs [8,11]. It is important to consider the data through this lens to understand the potential sources of bias and generalizability of our findings.

Many reviews explicitly compared an app to past experiences, in which another app did not meet the reviewer's needs. However, reviews of the included apps were generally positive, suggesting that users less frequently provided negative reviews when an app did not meet their needs. As a result, the content reviewed here may reflect a bias toward positive experiences. In addition, the duration of app use was unclear in most reviews. Future work should explore the features that yield positive first impressions and those associated with long-term app adherence.

It is also impossible to directly assess the health, digital, or data literacy of the reviewers. However, to generate the included content, users must have sufficient literacy to identify, download, use, and review health apps on a smartphone. Therefore, we presume that digital and health literacy in this population is moderate to high. Many reviewers requested sophisticated reports and visualizations or wished to export and analyze their data independently. This exceeds the expected data literacy of the general population [65], indicating a selection bias. Therefore, the results should be interpreted with caution in populations with low health, digital, and data literacy.

Conclusions

Data visualizations support depression self-management when they align with the users' individual needs and goals. To achieve this alignment, personalized data entry mechanisms and visualization content are often desired or necessary. These heterogeneous preferences pose a challenge for app developers, and further research should prioritize features based on their importance and impact on service users. Despite the limitations of the review-based content analysis, it contains readily attainable, free, and externally valid insights that complement formal qualitative research.

Acknowledgments

This study was written in support of the Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) program, a collaborative research effort focusing on the development of remote monitoring technology for monitoring central nervous system disease progression. We acknowledge all partners in the RADAR-CNS consortium [66] for their input and support in this work. The RADAR-CNS project received funding from the Innovative Medicines Initiative 2 Joint Undertaking [67] under grant agreement 115902. This joint undertaking receives support from the European Union's Horizon 2020 Research and Innovation Program and European Federation of Pharmaceutical Industries and Associations [67]. This communication reflects the views of the RADAR-CNS consortium, and neither the Innovative Medicines Initiative nor the European Union and European Federation of Pharmaceutical Industries and Associations are liable for any use that may be made of the information contained herein. This paper represents an independent research funded by the National Institute for Health Research Biomedical Research Centre in South London and Maudsley National Health Services Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the National Health Services, National Institute for Health Research, or Department of Health and Social Care. TW is a National Institute of Health Research senior investigator.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and tables describing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram, coding frame, and illustrative quotes.
[DOCX File , 841 KB - humanfactors_v9i4e40133_app1.docx]

Multimedia Appendix 2
Supplementary tables describing the characteristics of included apps and their reviews.
[XLSX File (Microsoft Excel File), 24 KB - humanfactors_v9i4e40133_app2.xlsx]

References

1. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321 [FREE Full text] [doi: 10.2196/15321] [Medline: 32012079]
2. Myers A, Chesebrough L, Hu R, Turchioe MR, Pathak J, Creber RM. Evaluating commercially available mobile apps for depression self-management. AMIA Annu Symp Proc 2020 Jan 25;2020:906-914 [FREE Full text] [Medline: 33936466]
3. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
4. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. Health Inf Sci Syst 2015 Feb 24;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S1] [Medline: 26019809]
5. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
6. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
7. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
8. Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, et al. Data visualization for chronic neurological and mental health condition self-management: systematic review of user perspectives. JMIR Ment Health 2022 Apr 28;9(4):e25249 [FREE Full text] [doi: 10.2196/25249] [Medline: 35482368]
9. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
10. Henze N, Pielot M. App stores: external validity for mobile HCI. Interactions 2013 Mar;20(2):33-38. [doi: 10.1145/2427076.2427084]
11. Arcia A, Merrill JA, Bakken S. Consumer engagement and empowerment through visualization of consumer-generated health data. In: Edmunds M, Hass C, Holve E, editors. Consumer Informatics and Digital Health: Solutions for Health and Health Care. Cham, Switzerland: Springer International Publishing; 2019:183-203.

12. Wu DT, Chen AT, Manning JD, Levy-Fix G, Backonja U, Borland D, et al. Evaluating visual analytics for health informatics applications: a systematic review from the American Medical Informatics Association Visual Analytics Working Group Task Force on Evaluation. J Am Med Inform Assoc 2019 Apr 01;26(4):314-323 [FREE Full text] [doi: 10.1093/jamia/ocy190] [Medline: 30840080]
13. Khasnabish S, Burns Z, Couch M, Mullin M, Newmark R, Dykes PC. Best practices for data visualization: creating and evaluating a report for an evidence-based fall prevention program. J Am Med Inform Assoc 2020 Feb 01;27(2):308-314 [FREE Full text] [doi: 10.1093/jamia/ocz190] [Medline: 31697326]
14. Wasil AR, Palermo EH, Lorenzo-Luaces LL, DeRubeis RJ. Is there an app for that? A review of popular apps for depression, anxiety, and well-being. Cogn Behav Pract 2022 Nov;29(4):883-901. [doi: 10.1016/j.cbpra.2021.07.001]
15. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
16. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
17. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
18. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
19. Rossi MG, Bigi S. mHealth for diabetes support: a systematic review of apps available on the Italian market. Mhealth 2017 May 4;3:16 [FREE Full text] [doi: 10.21037/mhealth.2017.04.06] [Medline: 28567412]
20. NHS Apps Library. NHS Digital. URL: https://digital.nhs.uk/services/nhs-apps-library [accessed 2020-07-31]
21. ORCHA. URL: https://appfinder.orcha.co.uk/ [accessed 2022-01-31]
22. Systematic Reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination. 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [accessed 2022-05-01]
23. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [doi: 10.1037/h0026256] [Medline: 19673146]
24. Ritchie J, Lewis J, McNaughton Nicholls C, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 2013.
25. 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 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
26. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
29. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
30. Fugard AJ, Potts HW. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. Int J Soc Res Methodol 2015 Feb 10;18(6):669-684. [doi: 10.1080/13645579.2015.1005453]
31. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb;18(1):59-82. [doi: 10.1177/1525822x05279903]
32. Loria S. TextBlob: Simplified Text Processing - Release v0.16.0. TextBlob. 2020. URL: https://textblob.readthedocs.io/en/dev/ [accessed 2022-05-01]
33. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inf Retr 2008;2(1-2):1-135. [doi: 10.1561/1500000011]
34. Anderson NH. Likableness ratings of 555 personality-trait words. J Pers Soc Psychol 1968 Jul;9(3):272-279. [doi: 10.1037/h0025907] [Medline: 5666976]
35. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927x09351676]
36. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Proc Int AAI Conf Web Soc Media 2021 Aug 3;7(1):128-137 [FREE Full text]
37. Park M, Cha C, Cha M. Depressive moods of users portrayed in Twitter. In: Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining. 2012 Presented at: SIGKDD '12; August 12-16, 2012; Beijing, China p. 1-8.

38. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semantics 2017 Mar 03;8(1):9 [FREE Full text] [doi: 10.1186/s13326-017-0120-6] [Medline: 28253919]
39. Jilka S, Odoi CM, van Bilsen J, Morris D, Erturk S, Cummins N, et al. Identifying schizophrenia stigma on Twitter: a proof of principle model using service user supervised machine learning. Schizophrenia (Heidelb) 2022 Feb 07;8(1):1 [FREE Full text] [doi: 10.1038/s41537-021-00197-6] [Medline: 35132080]
40. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965 Dec 01;52(3-4):591-611. [doi: 10.1093/biomet/52.3-4.591]
41. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
42. Fay MP, Proschan MA. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat Surv 2010;4:1-39 [FREE Full text] [doi: 10.1214/09-SS051] [Medline: 20414472]
43. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bull 1945 Dec;1(6):80-83. [doi: 10.2307/3001968]
44. Christensen RH. A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package. CRAN R Project. 2019 Dec 15. URL: https://cran.r-project.org/web/packages/ordinal/vignettes/clmm2_tutorial.pdf [accessed 2022-02-15]
45. Bolker B, Mächler M, Bates D, Scheipl F, Sørensen Ø, Green P, et al. lme4: Mixed-effects models in R. CRAN R Project. 2020. URL: https://github.com/lme4/lme4/ [accessed 2022-02-15]
46. R Core Team. The R Stats Package. Version 4.3.0. CRAN R Project. 2022. URL: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html [accessed 2022-11-15]
47. Fisher RA. The logic of inductive inference. J Royal Stat Soc 1935;98(1):39-82. [doi: 10.2307/2342435]
48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
49. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2021. URL: https://www.r-project.org/ [accessed 2022-02-15]
50. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
51. Matthews M, Voida S, Abdullah S, Doherty G, Choudhury T, Im S, et al. In situ design for mental illness: considering the pathology of bipolar disorder in mHealth design. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2015 Presented at: MobileHCI '15; August 24-27, 2015; Copenhagen, Denmark p. 86-97. [doi: 10.1145/2785830.2785866]
52. Matthews M, Abdullah S, Murnane E, Voida S, Choudhury T, Gay G, et al. Development and evaluation of a smartphone-based measure of social rhythms for bipolar disorder. Assessment 2016 Aug;23(4):472-483 [FREE Full text] [doi: 10.1177/1073191116656794] [Medline: 27358214]
53. Rohani DA, Tuxen N, Lopategui AQ, Faurholt-Jepsen M, Kessing LV, Bardram JE. Personalizing mental health: a feasibility study of a mobile behavioral activation tool for depressed patients. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 Presented at: PervasiveHealth '19; May 20-23, 2019; Trento, Italy p. 282-291. [doi: 10.1145/3329189.3329214]
54. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; April 27-May 2, 2013; Paris, France p. 2627-2636. [doi: 10.1145/2470654.2481364]
55. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
56. McClelland GT, Fitzgerald M. A participatory mobile application (app) development project with mental health service users and clinicians. Health Educ J 2018 Jun 05;77(7):815-827. [doi: 10.1177/0017896918773790]
57. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
58. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
59. Neuhauser L. Integrating participatory design and health literacy to improve research and interventions. Inf Serv Use 2017 Jun 26;37(2):153-176. [doi: 10.3233/isu-170829]
60. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly 2018 Jan 16;148:w14571 [FREE Full text] [doi: 10.4414/smw.2018.14571] [Medline: 29376547]
61. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: the 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]

62. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
63. Yarrington JS, Lasser J, Garcia D, Vargas JH, Couto DD, Marafon T, et al. Impact of the COVID-19 pandemic on mental health among 157,213 Americans. J Affect Disord 2021 May 01;286:64-70. [doi: 10.1016/j.jad.2021.02.056] [Medline: 33677184]
64. Strudwick G, Sockalingam S, Kassam I, Sequeira L, Bonato S, Youssef A, et al. Digital interventions to support population mental health in Canada during the COVID-19 pandemic: rapid review. JMIR Ment Health 2021 Mar 02;8(3):e26550 [FREE Full text] [doi: 10.2196/26550] [Medline: 33650985]
65. Turchioe MR, Myers A, Isaac S, Baik D, Grossman LV, Ancker JS, et al. A systematic review of patient-facing visualizations of personal health data. Appl Clin Inform 2019 Aug;10(4):751-770 [FREE Full text] [doi: 10.1055/s-0039-1697592] [Medline: 31597182]
66. RADAR-CNS: Remote Assessment of Disease and Relapse – Central Nervous System. RADAR-CNS. URL: https://www.radar-cns.org/ [accessed 2022-11-16]
67. IMI Innovative Medicines Initiative. URL: https://www.imi.europa.eu/ [accessed 2022-11-17]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Kushniruk; submitted 07.06.22; peer-reviewed by D Dinkel, E Pan, S Kandaswamy; comments to author 28.06.22; revised version received 19.07.22; accepted 06.08.22; published 23.11.22.

Please cite as:
Polhemus A, Simblett S, Dawe-Lane E, Gilpin G, Elliott B, Jilka S, Novak J, Nica RI, Temesi G, Wykes T
Health Tracking via Mobile Apps for Depression Self-management: Qualitative Content Analysis of User Reviews
JMIR Hum Factors 2022;9(4):e40133
URL: https://humanfactors.jmir.org/2022/4/e40133
doi:10.2196/40133
PMID:36416875

Original Paper

Public Trust in Artificial Intelligence Applications in Mental Health Care: Topic Modeling Analysis

Yi Shan[1], Prof Dr; Meng Ji[2], PhD; Wenxiu Xie[3], MPhil; Kam-Yiu Lam[3], PhD; Chi-Yin Chow[3], PhD

[1]Nantong University, Nantong, China

[2]University of Sydney, Sydney, Australia

[3]City University of Hong Kong, Hong Kong, China

Corresponding Author:
Yi Shan, Prof Dr
Nantong University
No 9, Seyuan Rd
Nantong, 226019
China
Phone: 86 15558121896
Email: victorsyhz@hotmail.com

Abstract

Background: Mental disorders (MDs) impose heavy burdens on health care (HC) systems and affect a growing number of people worldwide. The use of mobile health (mHealth) apps empowered by artificial intelligence (AI) is increasingly being resorted to as a possible solution.

Objective: This study adopted a topic modeling (TM) approach to investigate the public trust in AI apps in mental health care (MHC) by identifying the dominant topics and themes in user reviews of the 8 most relevant mental health (MH) apps with the largest numbers of reviewers.

Methods: We searched Google Play for the top MH apps with the largest numbers of reviewers, from which we selected the most relevant apps. Subsequently, we extracted data from user reviews posted from January 1, 2020, to April 2, 2022. After cleaning the extracted data using the Python text processing tool spaCy, we ascertained the optimal number of topics, drawing on the coherence scores and used latent Dirichlet allocation (LDA) TM to generate the most salient topics and related terms. We then classified the ascertained topics into different theme categories by plotting them onto a 2D plane via multidimensional scaling using the pyLDAvis visualization tool. Finally, we analyzed these topics and themes qualitatively to better understand the status of public trust in AI apps in MHC.

Results: From the top 20 MH apps with the largest numbers of reviewers retrieved, we chose the 8 (40%) most relevant apps: (1) Wysa: Anxiety Therapy Chatbot; (2) Youper Therapy; (3) MindDoc: Your Companion; (4) TalkLife for Anxiety, Depression & Stress; (5) 7 Cups: Online Therapy for Mental Health & Anxiety; (6) BetterHelp-Therapy; (7) Sanvello; and (8) InnerHour. These apps provided 14.2% (n=559), 11.0% (n=431), 13.7% (n=538), 8.8% (n=356), 14.1% (n=554), 11.9% (n=468), 9.2% (n=362), and 16.9% (n=663) of the collected 3931 reviews, respectively. The 4 dominant topics were topic 4 (cheering people up; n=1069, 27%), topic 3 (calming people down; n=1029, 26%), topic 2 (helping figure out the inner world; n=963, 25%), and topic 1 (being an alternative or complement to a therapist; n=870, 22%). Based on topic coherence and intertopic distance, topics 3 and 4 were combined into theme 3 (dispelling negative emotions), while topics 2 and 1 remained 2 separate themes: theme 2 (helping figure out the inner world) and theme 1 (being an alternative or complement to a therapist), respectively. These themes and topics, though involving some dissenting voices, reflected an overall high status of trust in AI apps.

Conclusions: This is the first study to investigate the public trust in AI apps in MHC from the perspective of user reviews using the TM technique. The automatic text analysis and complementary manual interpretation of the collected data allowed us to discover the dominant topics hidden in a data set and categorize these topics into different themes to reveal an overall high degree of public trust. The dissenting voices from users, though only a few, can serve as indicators for health providers and app developers to jointly improve these apps, which will ultimately facilitate the treatment of prevalent MDs and alleviate the overburdened HC systems worldwide.

(JMIR Hum Factors 2022;9(4):e38799) doi:10.2196/38799

KEYWORDS

public trust; public opinion; AI application; artificial intelligence; mental health care; topic modeling; topic; theme; term; visualization; user feedback; user review; Google Play; health app: mHealth; mobile health; digital health; eHealth; mental health; mental illness; mental disorder

Introduction

Background

"Mental disorders are one of the greatest public health concerns of our time" [1]. It is estimated that mental disorders (MDs) account for 32.4% of years lived with disability and 13.0% of disability-adjusted life years [2]. MDs have a lifetime prevalence ranging from 12% to 47.4% worldwide [3]. They affect millions of people worldwide, imposing a heavy burden on health care (HC) systems [4]. The heavy burden of MDs calls for increasing mental health (MH) research worldwide [1].

A possible resolution to this burden is adopting mobile health (mHealth) technologies, especially mobile phone apps [4]. Mental health care (MHC) is "having a digital moment" [5]. Digital mental health care (DMHC) takes diverse forms, varying from online psychotherapy (eg, cognitive behavioral therapy) to chatbots, which are computer programs using artificial intelligence (AI) and natural language processing to engage in conversations with text or text to speech [6]. These AI apps in MHC boast unmatched advantages, including "increased convenience for patients, enhanced adherence to appointments, and access to care that is unbound by geography, backed by evidence that digital mental health care can be effective" [5]. Providing evidence-based care, AI health apps have become increasingly popular, evidenced by over 150,000 downloads of an app named PTSD Coach [7]. However, if not guided by human therapists, they have high dropout rates [8]. By investigating the most popular MH apps downloaded over 100,000 times, Baumel et al [9] discovered that 96% of users no longer engaged with the apps after 2 weeks. This is because DMHC may not guarantee "patient privacy, confidentiality, and reliability of service delivery" [5]. According to previous studies, what influences the implementation of AI apps includes (1) people-related factors (eg, public attitudes, trust), (2) health system–related factors (eg, clinical responsibility and accountability, the possibility of harm, and issues of regulation and service provision), (3) data-related factors (eg, issues of data security, privacy, consent, and ownership), and (4) tool-related factors (eg, issues of reliability and validity) [10-12].

"[The] COVID-19 pandemic has changed health care delivery, including mental health care services around the world" [5]. DMHC has been "a welcome and much-needed adoption" in the current pandemic, mainly because no safe alternative can provide MH services. DMHC helps to sustain core MH services. In this state of research, to what extent do users trust AI apps in MHC?

Trust Defined and Trust in AI Systems (Apps)

As an important social lubricant for cooperative behavior [13], trust has been investigated in various disciplines, which have provided varying definitions, leading to "a multidimensional family of trust concepts," each with a specific focus [13]. Trust is fundamentally "a feeling of certainty that a person or a thing will not fail," often based on inconclusive evidence and categorized into interpersonal trust, social trust, and trust in automation [14]. It is 1 of the means that people can use to reduce complexity in a complex world by decreasing the number of choices to be considered in particular circumstances [15,16]. Since trust can reduce fear, risk, and complexity both online and offline, it is inconceivable that a robust, interactive online environment would be possible without trust [13].

The concept of human trust has been investigated in automated systems [17] and online systems [13]. Reeves and Nass [18] studied how people treated new technologies as humans and as objects of trust, finding that people responded to these technologies almost in the same way that they responded to real people in social relationships by behaving politely or rudely to computer systems; regarding them as assertive, timid, or helpful agents; and responding to them physically. Trust in technology is the belief that a tool, machine, or equipment will not fail [19]. "Trusting relationships with the technology have the potential to affect the way the technology is used or not used" [16]. The literature on trust in HC systems focused on patient-physician interpersonal trust [12,20,21] and patients' trust in HC systems [22,23]. Most research on trust in technology pointed to the trust of the technology operators, for example, the care providers (physicians, nurses, or technicians) in medical settings [24]. However, trust in technology needs to be studied from the perspective of users. It is their trust in automated systems that results in appropriate use, disuse, misuse, or even abuse of the automation [25].

Trust in information sources plays an essential part in making the individual well informed of information and willing to act upon it [26]. For example, people perceive their risks of COVID-19 infection differently, depending on whether the information is received from social media or from mass media [27]. It has been found that messages from privately affiliated media sources can undermine people's trust in scientific knowledge and health policies [28]. Trust is an essential factor that impacts human interactions with AI [29]. It is crucial to understand human-AI trust dynamics, especially in the HC domain, in which life may be at risk [29]. User perception of AI systems' capabilities is always a significant factor for trust in them [29], and the degree of trust in them considerably influences the level of user reliance on them [30] and thus the user efficacy of HC decisions [29]. Reliability (ie, AI systems' predictable and consistent performance of tasks [31]) is particularly concerning in HC because AI changes in reliability in the presence of new data [32]. The reliability of AI systems is also conditioned both by input data and by user data [29]. When trained with insufficient and subjective data from diverse sources, AI systems might generate overfitted or even biased outcomes of which clinical users could not be aware [29]. These factors undermine the performance of AI technology [33], preventing users from trusting and accepting AI systems [29].

Understanding trust in AI systems will shed light on decision-making concerning the acceptance or rejection of technologies, the system designs bringing about positive patient outcomes, and those imposing adverse effects [14].

To the best of our knowledge based on literature search and review, no study has exclusively investigated public trust in AI apps or systems in MHC, although some studies have examined trust in medical technology [14]; human trust in AI in HC [29]; public trust in COVID-19–related government information sources (eg, the US Centers for Disease Control and Prevention), private sources (eg, FOX and CNN), and social networks (eg, Facebook and Twitter) [34]; and public trust in COVID-19–tracing apps [35]. In this state of research, there is a pressing need to explore public trust in AI apps or systems in MHC, particularly in the current context where MDs are listed among the greatest public health concerns of our times [1].

Objective

This study aims to investigate public trust in AI apps in MHC by identifying the major topics (themes) in users' reviews of the 8 most relevant MH apps with the largest numbers of reviewers.

Based on this research objective, we proposed the following research questions:

- Research question 1: Did the users trust the 8 most reviewed MH apps?
- Research question 2: What dominant topics (themes) and most relevant terms could be identified in user reviews?
- Research question 3: Were the identified dominant topics (themes) concerned with public trust or mistrust?
- Research question 4: What implications can the identified trust or mistrust provide for developers and providers of MH apps?

Topic modeling (TM) was adopted to process the reviews of these apps. As a statistical model, TM arranges unstructured data structurally using latent themes [36]. With this model, we scrutinized the reviews of the apps under discussion to reveal the extent to which users trusted these apps. The findings could not only fill the gap in the literature but also facilitate the cooperation between HC practitioners and app developers to design MH intervention programs for these MH apps to respond to public concerns effectively.

Methods

Data Collection

We searched Google Play for the top 20 MH apps with the largest numbers of reviewers, from which we selected the 8 (40%) most relevant apps: (1) *Wysa: Anxiety Therapy Chatbot*; (2) *Youper Therapy*; (3) *MindDoc: Your Companion*; (4) *TalkLife for Anxiety, Depression & Stress*; (5) *7 Cups: Online Therapy for Mental Health & Anxiety*; (6) *BetterHelp-Therapy*; (7) *Sanvello*; and (8) *InnerHour*. Subsequently, we extracted the data of user reviews of these apps posted from January 1, 2020, to April 2, 2022, by using a Python web crawler. After extracting the data, we created a data set of 3931 reviews (reviews of less than 20 words were removed) that consisted of the names of the selected apps, the content of these reviews, the date when these reviews were published, the dominant topic in each review, the contribution of the dominant topic to each review measured in percentage points, and the terms related to each topic. The data set is given in Multimedia Appendix 1. Before the analysis, we followed the standard preprocessing procedures designed in previous studies [37,38] to clean the data using Python 3.0 (Python Software Foundation) and to perform word part-of-speech tagging and text processing using the Python library *spaCy* [39,40]. Through data cleaning, we converted the words in the reviews into lowercase words; removed stop words, punctuation, numbers, and nonword characters; and stemmed the remaining text [41]. To generate more interpretable topics of high quality, we restricted the parts of speech of words to "noun" (NOUN), "verb" (VERB), "adjective" (ADJ), or "proper noun" (PROPN). The standard preprocessing procedures can significantly enhance the performance of algorithms and stabilize the stochastic inference of latent Dirichlet allocation (LDA) [38].

Topic Modeling With LDA

The statistical methods of unsupervised TM algorithms (which do not need prior labeling or annotations of the documents) were designed to analyze the words (terms) of the original texts to identify the themes (topics) running through a corpus [42,43]. These algorithms allow users to organize and summarize numerous documents that cannot be annotated manually [41], thereby revealing the hidden topics in the documents [43]. We adopted the LDA TM technique, which assumes that texts are generated from a mixture of topics [44]. LDA is efficient and can generate topics of better quality [45]. From the data set created, we generated 2 probability distribution outputs: the probability distribution of topics over documents and the probability distribution of terms over topics [41,43]. The number of topics was determined by repeating the analysis with different numbers of topics and by comparing the perplexity of each analysis [41]. A lower perplexity value indicates a better model fit [44], and the perplexity value decreases with the increase in the number of topics [41]. Both the simplicity and the interpretability of the textual content need to be considered in choosing the optimal number of topics [38].

We took the coherence score as an assessment metric to evaluate how good a given topic model was and determine the optimal number of topics [46] that needed to be extracted from the user reviews. Topic coherence is a qualitative method used to score the coherence of a given topic [46]. It measures the semantic similarity between words with high scores in a topic to determine the consistency of a single topic, improving the semantic understanding of the topic [36]. We applied the Python package coherence model from *Gensim* to calculate the coherence value [47]. As shown in Figure 1, the coherence score increased to the highest value of 0.45 when the number of topics reached 4, before decreasing gradually, implying that the optimal number of topics was 4. Afterward, we visualized the relationship between these 4 topics and their related terms using Python version 3.6.1 and the *pyLDAvis* tool [44].

When λ equals 1, terms are sorted according to their frequency in a topic. Therefore, we set λ=1 to visualize the intertopic

distance between the 4 topics and the top 30 most relevant terms for each topic, as shown in Figure 2. We classified these 4 topics into different themes to facilitate better analysis based on the computed topic distance [44]. In the 2D plane (Figure 2), the 4 topics are shown in the form of 4 circles. The size of each circle represents the overall prevalence of the topic, the overlap between circles 3 and 4 means the overlap between topics 3 and 4, and the distance between the circle centers stands for topic distance [44]. The content of each topic was generated according to its corresponding set of keywords (terms) [48]. Considering that the output of statistical measures cannot be guaranteed to be interpretable due to the language complexity [49], we complemented automatic text statistics with manual interpretation when analyzing the topics. The topics were named based on the associated keywords to illustrate those topics [48].

Figure 1. Coherence score for the topic numbers.

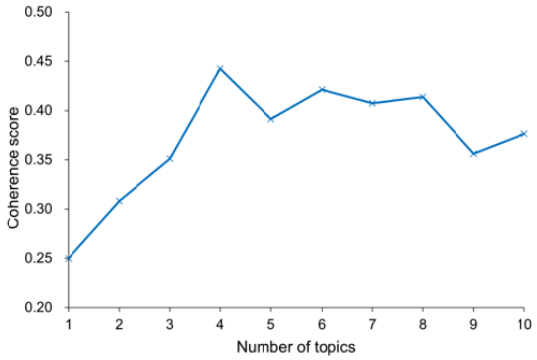
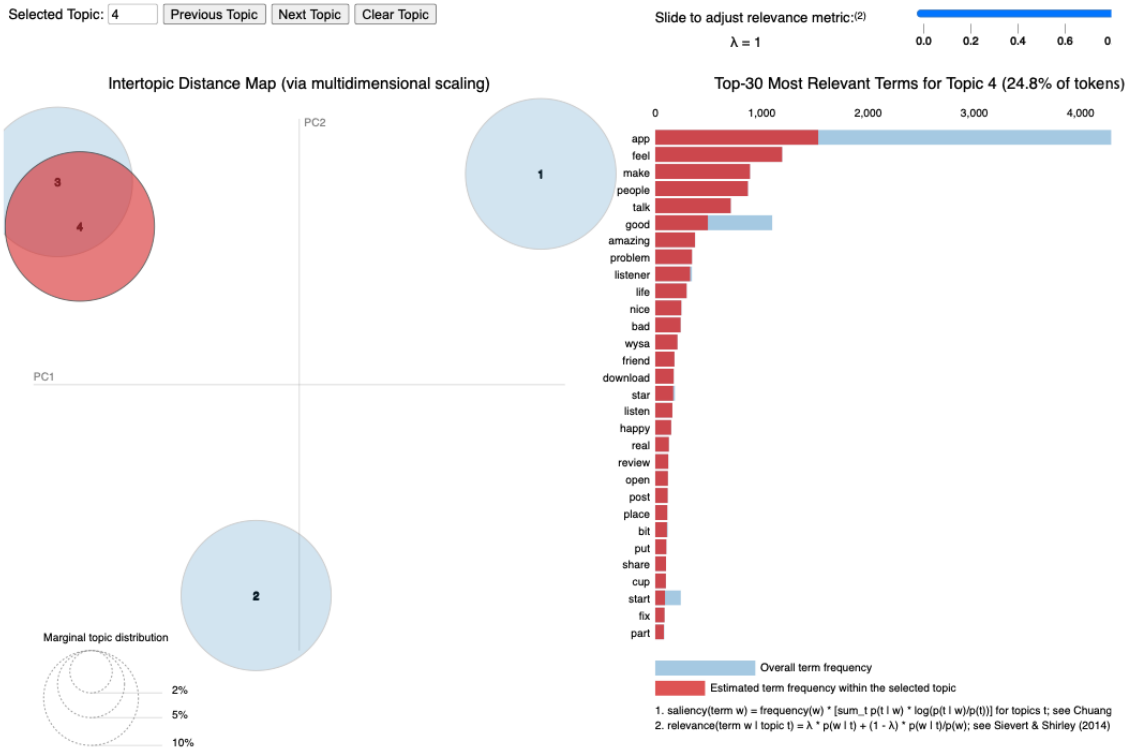


Figure 2. Intertopic distance map and top 30 most relevant terms for topic 4. Refer to the interactive web-based visualization in Multimedia Appendix 1 for other topics. PC: principal component.



Ethical Considerations

All reviewers anonymized themselves when posting app reviews online. We just downloaded the anonymized data from the apps. Therefore, an ethical review was not necessary for this paper.

Results

Review Distribution Over Apps and Time

Figure 3 shows the distribution of reviews over the 8 selected apps. Over the period from January 1, 2020, to April 2, 2022, the largest number of reviews were posted on *InnerHour* (n=663, 16.9%), followed by *Wysa: Anxiety Therapy Chatbot* (n=559, 14.2%), *MindDoc: Your Companion* (n=538, 13.7%), and *7 Cups: Online Therapy for Mental Health & Anxiety* (n=554, 14.1%). Less than 450 reviews were published on *Youper Therapy* (n=431, 11%) and *BetterHelp-Therapy* (n=468, 11.9%). Far fewer than 400 reviews were released on *TalkLife for Anxiety* (n=356, 8.8%) and *Sanvello* (n=362, 9.2%). The different numbers of reviews somehow reflected the varying popularity of these 8 apps over the period under discussion. Figure 4 displays the number of reviews published over time,

which peaked and waned on different dates. Although it is difficult to attribute the 2 peaks in 2021 to specific factors, it is likely that the peaking period in the middle of 2020 possibly reflected the increased need for MHC apps during the global outbreak of the COVID-19 pandemic, as evidenced by the following reviews:

> *It's really helpful specially during this pandemic. You can find very calm listeners who will be patient and understanding while listening.*

> *Love this app. Tried it because of a friend's recommendation and best decision in this challenging time of a pandemic. Really cleared my mind and helped me regain control and life perspective. Will make it a daily habit.*

> *This app is perfect for these times. With anxiety issues predating the pandemic, I'll probably continue using it after the world gets back on track. Thank you!*

Figure 3. Number of reviews by app.

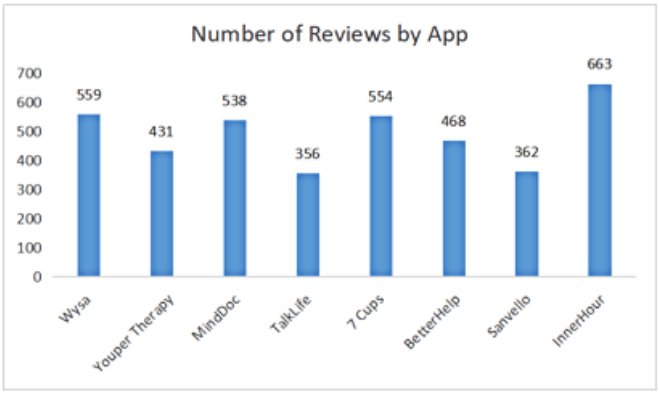
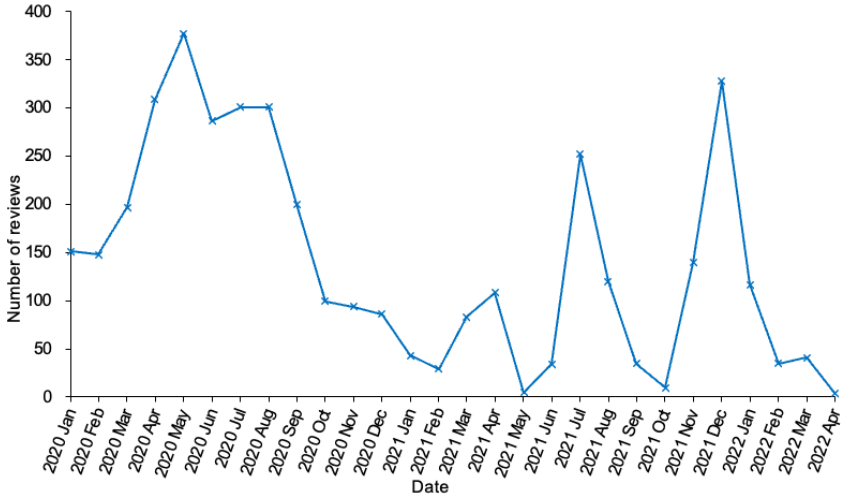


Figure 4. Number of reviews over time.



Topic Prevalence

By applying LDA TM, we classified into 4 topics the collected data of user reviews of the 8 selected AI apps in MHC published from January 1, 2020, to April 2, 2022 (Table 1). Topic 4 (*cheering people up*) was the most popular and dominant topic, occurring in 27.2% (n=1069) of the 3931 reviews collected, closely followed by topic 3 (*calming people down*), which appeared in 26.2% (n=1029) of the 3931 reviews. Topic 2 (*helping figure out the inner world*) ranked as the third-most popular and dominant topic, showing up in 24.5% (n=963) of the 3931 reviews. Topic 1 (*being an alternative or complement to a therapist*) accounted for 22.1% (n=870) of the 3931 reviews. The similar areas of the 4 circles representing these 4 topics in Figure 2 indicate the similar overall prevalence of these topics in the collected reviews.

Table 1. Topic classification and keywords.

Theme and classification	Keywords	Reviews (N=3931), n (%)
Theme 1: being an alternative or complement to a therapist		
Topic 1: being an alternative or complement to a therapist	great, therapist, time, work, therapy, issue, chat, counselor, easy, option, week, person, experience, money, support, month, session, message, year, service, start, worth, account, enjoy, video, conversation, update, access, call, log	870 (22.1)
Theme 2: helping figure out the inner world		
Topic 2: helping figure out the inner world	helpful, day, give, find, good, app, mood, time, question, feeling, mental_health, track, thought, change, emotion, answer, phone, notification, check, add, daily, tool, positive, community, hard, journal, provide, insight, negative, advice	963 (24.5)
Theme 3: dispelling negative emotions		
Topic 3: calming people down	app, love, thing, pay, lot, free, anxiety, recommend, depression, premium, feature, understand, stress, calm, AI[a], version, hope, struggle, care, exercise, meditation, sleep, afford, activity, improve, good, offer, a lot, awesome, deal	1029 (26.2)
Topic 4: cheering people up	app, feel, make, people, talk, good, amazing, problem, listener, life, nice, bad, Wysa, friend, download, star, listen, happy, real, review, open, post, place, bit, put, share, cup, start, fix, part	1069 (27.2)

[a]AI: artificial intelligence.

Intertopic Distance

Figure 2 provides an overview of the topic model we constructed. The 4 circles in this figure represent the 4 topics dominating the 3931 reviews. The intertopic distances characterized by multidimensional scaling on the 2D plane in Figure 2 imply the semantic similarity between these 4 topics: topics 3 and 4 overlap and are thus semantically similar; these 2 topics are semantically distant from topics 1 and 2, which are also semantically distant from each other.

Topic Terms and Content

Figure 2 also shows the top 30 most relevant terms for topic 4, representing 24.8% of the tokens. As this topic constitutes the highest proportion (n=1069, 27.2%) of the collected reviews, it is presented as a typical illustration. Since the blue bar and the red bar represent the overall term frequency and the estimated term frequency within the selected topic, respectively, the topic content can be better interpreted based on this approach [50,51]. With regard to topic 4, users of the 8 selected AI apps in MHC preferred to use words indicating the apps' function of cheering people up, such as *make*, *feel*, *happy*, *good*, *amazing*, *listener*, *friend*, and *nice*. In this way, we could study the content of this topic and name it. The word clouds of the 4 dominant topics in Multimedia Appendix 2 complement Figure 2 in demonstrating the term frequency of the top 30 most relevant terms of each topic and implying the content of each topic.

Topic Classification and Keywords

Drawing on TM to analyze the user reviews and compare the perplexity indices, we discovered that it was optimal to classify the 4 dominant topics covered in the reviews into 3 themes, as shown in Table 1. The illustrative quotes for these themes are displayed in Textbox 1.

Topic 1 (theme 1, "being an alternative or complement to a therapist") deals with reviews that describe the apps' function as an alternative or even a complement to a human MH provider (eg, nurse, doctor, physician, psychiatrist) when such a human therapist is unavailable due to the limitation of working time, space, budget, and social conditions (eg, the repeated resurgences of COVID-19); when app users attach great importance to privacy and social stigma related to MH problems; or when DMHC takes prevalence or will be institutionalized in the MHC domain. Topic 2 (theme 2, "helping figure out the inner world") focuses on the apps' role in helping users track their psychic world to gain a better understanding of their feelings and emotions. By expressing every emotional state to the apps, users can keep themselves in control emotionally, turning out to be a better, sounder self mentally. Topics 3 and 4 (theme 3, "dispelling negative emotions") point to the emotion-soothing purpose of the apps under discussion. Specifically, these apps can serve to mitigate stress, anxiety, and depression that haunt users, effectively calming them down and cheering them up by behaving as a bosom friend, a good listener, and a constructive adviser, while avoiding judging them by what they say during their venting of various types of feelings.

Textbox 1. Illustrative quote(s) of each theme.

Theme 1

- "Great! Very calming and I haven't been able to see my therapist in a while and this really helped. Great quality too."

Theme 2

- "This app has helped me a lot on my mental health. I feel much more in tune with myself and my emotions and all because this app helped me figure out myself and my inner world better. Plus I also have someone I can express everything at the end of the day."

Theme 3

- "It's a great help. It calms me down some and it's calming to know that I wouldn't be judged by what I say."
- "Thanks to the Wysa team for making this! It's been making it much easier for me to manage my stress (:"

Discussion

Principal Findings

MDs represent a tremendous public health concern of our times [1], which is increasingly overburdening the HC systems worldwide. This is especially true during the COVID-19 pandemic, which has caused many MDs worldwide [52]. In this context, Web 2.0–empowered DMHC apps are an optimal solution because there is currently a "tectonic shift in the ways in which patients consume health and medical information" [53]. Digital health interventions have proved effective in mitigating negative health habits and outcomes, leveraging patients' growing proactivity in obtaining health information online [54]. This study is the first investigation of public trust in AI apps in MHC from the perspective of user reviews of MHC apps. Applying TM, we managed to manipulate unstructured user reviews, discovering 4 dominant topics hidden in a data set of 3931 reviews and categorizing these topics into 3 themes based on topic coherence and intertopic distance (semantic similarity).

The top 30 most relevant terms for the 4 dominant topics (Figures 2 and Multimedia Appendix 1) indicate that users of the 8 selected apps generally trust these apps, having a positive experience with these apps and feeling satisfied with them. Specifically, this was evidenced by the top terms *great*, *therapist*, *work*, *chat*, *counselor*, *easy*, and *option* in topic 1; *helpful*, *find*, *give*, *mood*, *good*, *question*, *positive*, and *feeling* in topic 2; *app*, *love*, *free*, *pay*, *lot*, *recommend*, *anxiety*, *depression*, and *stress* in topic 3; and *app*, *make*, *feel*, *people*, *talk*, *good*, *life*, *amazing*, *listener*, *nice*, *friend*, *open*, and *happy* in topic 4. In medical settings, trust results in enhanced health self-efficacy, increased treatment adherence, and more positive health outcomes [55,56]. This result confirms a previous research finding that overall user trust in conversational agents (CAs) is increasing [57].

The 4 topics that the users focused on were classified into 3 themes based on LDA TM. Theme 3 (*dispelling negative emotions*) covered topics 3 (*calming people down*) and 4 (*cheering people up*), which together dominated more than half (n=2098, 53.4%) of the reviews (N=3931) under scrutiny. This theme highlighted the apps' perceived function of helping people out of depression, anxiety, and stress. The results revealed that the users are mainly concerned with positive MH outcomes when adopting AI apps for MH interventions. They sought social support messages concerning MH on MHC apps mainly because they trusted the apps in terms of (1) the look of the technology, (2) perceived reliability, (3) accuracy, (4) consistency, and (5) feedback from the technology [14], as demonstrated by the following reviews:

> *The app is also easy to navigate and seems solid (no tech issues). Also, nice professional appearance.*
>
> *Very cute and user-friendly layout. Takes you step by step to teach you by example how to take the reins on your own thoughts and emotions. Awesome app!*
>
> *Reliable. It's a really great app, the listeners out on the platform are professional and patient. I have personally benefited from this app and I can't stress this enough, it is a really good app. If you just need an ear to listen to you, without any judgments whatsoever, try it.*
>
> *It's scary how meticulously accurate this personality test is. It is like having a paid friend Psychologist on hand during these uncertain times of total lockdown...*
>
> *7 cups has consistently and diligently kept the essence of being there for anyone anywhere anytime, very alive and accessible*
>
> *I love being able to message my counselor whenever I need to, it helps with needing to vent and get positive, constructive feedback quickly, on everyday stressors that feel overwhelming.*

These quoted reviews show that some reviewers tend to trust computer systems more than other humans (ie, human MHC providers) because automation is expected to be relatively perfect [58], in terms of appearance, reliability, accuracy, consistence, and targeted feedback [14].

Moreover, our results confirmed that *helping figure out the inner world* (theme 2) was another important factor determining public trust. Reviewers repeatedly mentioned *thought tracking*, *best mood tracker*, *journalizing my thoughts and feelings*, *making me take a proper look at my thoughts and feelings*, *writing down my thoughts and then framing them to be more positive*, etc. The function of enabling users to follow through with their MH made them trust their MH with these apps due to the negative social perception of MDs [59], which prevents people from seeking professional help proactively [60]. This result aligns with previous studies [4,59] that mHealth

technologies (eg, MH apps) can protect users' privacy and reduce the influence of the social stigma attached to MDs. Users commented that "it can be nice to vent to something that isn't someone who would judge you." In addition, people are most likely to depend on the advice and support provided by professional apps when developing appropriate responses to MH problems [54]. The professional interventions in thought tracking supplied by the selected apps were another essential contributor to the high degree of public trust in these apps, as evidenced by some reviews, such as:

> *The bot navigates you through series of practical and professional activities*
>
> *Very professional*
>
> *Professional help*
>
> *Helps me assemble my thoughts by asking me relevant questions*

Additionally, our research results suggest that MH app users also reflect on the role of MH apps as an alternative or a supplement to human therapists (theme 1). This implies that app users attach great importance to digital health interventions, while still trusting human MHC providers. Although some users accepted MH apps as an alternative to human therapists, for example,

> *This app is my perfect personal therapist,*

most users took them as a complement to human therapists, for example,

> *Very helpful between therapy appointments and during times when I wasn't seeing a therapist.*

The tectonic change in patients' consuming health and medical information [53] significantly shifted the landscape for interventions seeking to change health behavior, each with its associated benefits and risks [61]. "Although exposure to these products is becoming ubiquitous, electronic health information is novel, incompletely disseminated, and frequently inaccurate, which decreases public trust" [54]. That is why a few users mistrusted the selected AI apps in their reviews:

> *It's the same vague questions every time with no room for explanation. Very inaccurate*
>
> *But the therapists seem very disconnected and don't really seem to care much at all, responding in broken or incomplete sentences*

However, AI apps in MHC are gaining increasing popularity due to their unmatched advantages [5], the provision of evidence-based care [7], and the change in HC delivery due to the COVID-19 pandemic [5]. As a result, AI apps in MHC have become "a welcome and much-needed adoption" [5].

Admittedly, there were some dissenting voices in the collected reviews. These disagreements reflected some aspects of the selected apps that undermined public trust, including:

- Interface design

  The interface sucks

- Charges

  It was an exorbitant in-app purchase

- Engagement

  ...chat it was not engaged at all

- Reliability

  Broken app. Chats are so unstable I couldn't get more than 2 messages before there was an error. The interface is broken, half of the stuff barely works and everything lags. The idea behind it is amazing but in practice it unfortunately doesn't work at all.

- Ease of use

  The app itself isn't as easy to navigate compared to the desktop site

These negative comments of reviewers indicate that the reliability of service [5,29,31], AI systems' capabilities [29], people-related factors (eg, public attitudes, trust), health system–related factors (eg, issues of regulation and service provision), and tool-related factors (eg, issues of reliability) [10-12] all potentially reduce public trust in DMHC delivered through AI apps or systems. Among these factors damaging public trust, reliability or predictable and consistent performance arouses particular concern in HC [32]. These findings are consistent with those reported in previous studies [5,10-12,29,31,32].

All these aspects need to be improved to enhance public trust. The technology design and the integration of technology into effectively performing systems are most likely to be associated with patient trust, which impacts patient outcomes (eg, patient satisfaction and adherence to medical advice and treatment) [62].

Overall, the investigation of user reviews based on the TM approach can shed light on public trust in AI apps in MHC. "Understanding trust in medical technologies will provide insight into decision making about which technologies will be accepted or rejected, which work system designs will lead to positive patient outcomes, and which will have the inverse effect" [14].

Implications

Trust is an emotional construct that features patients' comfortable feeling of faith in or dependence on care providers' intentions, with common dimensions, including "competence, compassion, privacy and confidentiality, reliability and dependability, and communication" [63]. In the current and future context of technology replacing human elements in medical practices [14] and HC providers depending more on AI [29], a proper trust relationship [29] or calibrated trust [64] needs to be established between users and AI for HC for effective medical and health decisions [29]. To this end, important issues relevant to trust in AI for HC apps should be studied, including essential factors impacting trust in AI for HC, potential ways to improve trust relationships, and their influence on trust [29].

Currently, the adoption of AI systems in the HC domain is considerably hindered by a lack of trust in this technology [29]. Trust in AI is conditioned both by human factors and by properties of AI systems. Human factors include users'

education, experiences, personal biases, and perception of automation [29]. Properties of AI systems involve the look of the technology, perceived reliability, accuracy, consistency, and feedback from the technology [14]. Considering these contributors to public trust in AI systems, we propose that patients be provided with education and information about the technology and its use [14] and that developers of AI apps for HC well consider health system–related factors, data-related factors, and tool-related factors [10-12] when designing HC apps and systems. In addition, mechanisms need to be incorporated into the development of AI to build and keep an appropriately balanced, optimal level of user trust matching the capacities of AI systems [30].

Limitations

This study may have some limitations. Most importantly, user reviews of the selected apps were subjective user comments based on the individual experience with the apps. Public trust in AI apps in MHC reflected by such self-reported data may be influenced by various factors, such as prior experience with the source, sociodemographic background, or health behaviors [54]. Thus, the user reviews may be biased, possibly influencing our research findings to some extent. Second, the user experience and satisfaction with AI apps in MHC are likely to be impacted by the current social situations. Therefore, public trust in such apps may vary according to changing social conditions, such as the current resurgences of the COVID-19 pandemic. Such changing status of public trust may thus not necessarily be triggered by the improved or deteriorating performance of the AI apps in MHC. The overall high degree of public trust found in this study may attenuate when the pandemic ends. Future studies need to consider the dynamics of the review data. Third, this study only retrieved reviews posted from January 1, 2020, to April 2, 2022. Further studies may expand this period to test the generalizability of our findings. Fourth, future research needs to consider more apps to verify the generalizability of the results of this study.

Conclusion

This is the first study investigating the public trust in AI apps in MHC from the perspective of user reviews using the TM technique. The automatic text analysis and complementary manual interpretation of the collected reviews allowed us to discover 4 dominant topics hidden in a data set of 3931 unstructured reviews and categorize these topics into 3 groups of themes based on topic coherence and intertopic distance (semantic similarity). From these topics and themes, we managed to study the status of public trust in AI apps in MHC, finding an overall high degree of public trust. Although shaped by various factors, public trust in technology can broadly reflect the performance of the technology itself. The negative voices from users can serve as indicators for health providers and app developers to jointly improve these apps, which will ultimately facilitate the treatment of prevalent MDs and alleviate the overburdened HC systems worldwide. More research needs to be conducted into the design of AI apps for health counseling that ensures the validity of any provided recommendations and interventions. Meanwhile, patients must be reminded that health recommendations from any nonauthoritative sources need to be confirmed with HC professionals before they are acted on [65].

Conflicts of Interest

None declared.

Multimedia Appendix 1
App review data and topic modeling results.
[XLSX File (Microsoft Excel File), 541 KB - humanfactors_v9i4e38799_app1.xlsx]

Multimedia Appendix 2
Topic Modeling word clouds of the 4 dominant topics.
[PNG File , 258 KB - humanfactors_v9i4e38799_app2.png]

References

1. Motrico E, Salinas-Perez JA, Rodero-Cosano ML, Conejo-Cerón S. Editors' comments on the special issue "Social Determinants of Mental Health". Int J Environ Res Public Health 2021 Apr 09;18(8):3957 [FREE Full text] [doi: 10.3390/ijerph18083957] [Medline: 33918698]
2. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/s2215-0366(15)00505-2]
3. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
4. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR Mhealth Uhealth 2019 Nov 07;7(11):e13236 [FREE Full text] [doi: 10.2196/13236] [Medline: 31697245]
5. Gratzer D, Torous J, Lam RW, Patten SB, Kutcher S, Chan S, et al. Our digital moment: innovations and opportunities in digital mental health care. Can J Psychiatry 2021 Jan 30;66(1):5-8 [FREE Full text] [doi: 10.1177/0706743720937833] [Medline: 32603188]

6. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
7. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD Coach. JMIR Ment Health 2015 Mar;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
8. Gratzer D, Khalid-Khan F. Internet-delivered cognitive behavioural therapy in the treatment of psychiatric illness. CMAJ 2016 Mar 01;188(4):263-272 [FREE Full text] [doi: 10.1503/cmaj.150007] [Medline: 26527829]
9. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
10. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
11. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-e13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
12. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
13. Corritore CL, Kracher B, Wiedenbeck S. On-line trust: concepts, evolving themes, a model. Int J Hum-Comput Stud 2003 Jun;58(6):737-758. [doi: 10.1016/s1071-5819(03)00041-7]
14. Montague EN, Winchester WW, Kleiner BM. Trust in medical technology by patients and health care providers in obstetric work systems. Behav Inf Technol 2010 Sep;29(5):541-554 [FREE Full text] [doi: 10.1080/01449291003752914] [Medline: 20802836]
15. Barber B. The Logic and Limits of Trust. New Brunswick, NJ: Rutgers University Press; 1983.
16. Lewis JD, Weigert A. Trust as a social reality. Soc Forces 1985 Jun;63(4):967. [doi: 10.2307/2578601]
17. Madhavan P, Wiegmann DA, Lacson FC. Automation failures on tasks easily performed by operators undermine trust in automated aids. Hum Factors 2006 Sep 06;48(2):241-256. [doi: 10.1518/001872006777724408] [Medline: 16884046]
18. Computers & Mathematics with Applications. Stanford, CA: Center for the Study of Language and Information/Cambridge University Press; 1997:128.
19. Sheridan TB. Humans and Automation. Santa Monica, CA: John Wiley and Sons; 2002.
20. Tarn DM, Meredith LS, Kagawa-Singer M, Matsumura S, Bito S, Oye RK, et al. Trust in one's physician: the role of ethnic match, autonomy, acculturation, and religiosity among Japanese and Japanese Americans. Ann Fam Med 2005 Jul 01;3(4):339-347 [FREE Full text] [doi: 10.1370/afm.289] [Medline: 16046567]
21. Arora NK, Gustafson DH. Perceived helpfulness of physicians' communication behavior and breast cancer patients' level of trust over time. J Gen Intern Med 2009 Feb 17;24(2):252-255 [FREE Full text] [doi: 10.1007/s11606-008-0880-x] [Medline: 19089501]
22. Zheng B, Hall MA, Dugan E, Kidd KE, Levine D. Development of a scale to measure patients' trust in health insurers. Health Serv Res 2002 Feb;37(1):187-202. [Medline: 11949920]
23. Balkrishnan R, Hall MA, Blackwelder S, Bradley D. Trust in insurers and access to physicians: associated enrollee behaviors and changes over time. Health Serv Res 2004 Aug;39(4 Pt 1):813-823 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00259.x] [Medline: 15230929]
24. Parasuraman R, Miller CA. Trust and etiquette in high-criticality automated systems. Commun ACM 2004 Apr;47(4):51-55. [doi: 10.1145/975817.975844]
25. Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Hum Factors 2016 Nov 23;39(2):230-253. [doi: 10.1518/001872097778543886]
26. Flowers P, Riddell J, Boydell N, Teal G, Coia N, McDaid L. What are mass media interventions made of? Exploring the active content of interventions designed to increase HIV testing in gay men within a systematic review. Br J Health Psychol 2019 Sep 02;24(3):704-737 [FREE Full text] [doi: 10.1111/bjhp.12377] [Medline: 31267624]
27. Lin Y, Hu Z, Alias H, Wong LP. Influence of mass and social media on psychobehavioral responses among medical students during the downward trend of COVID-19 in Fujian, China: cross-sectional study. J Med Internet Res 2020 Jul 20;22(7):e19982 [FREE Full text] [doi: 10.2196/19982] [Medline: 32584779]
28. Simonov A, Sacher S, Dubé J, Biswas S. Frontiers: the persuasive effect of Fox News: non-compliance with social distancing during the COVID-19 pandemic. Mark Sci 2022 Mar 16;41(2):230-242. [doi: 10.1287/mksc.2021.1328]
29. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
30. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]

31. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans Manage Inf Syst 2011 Jun 01;2(2):1-25. [doi: 10.1145/1985347.1985353]
32. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science 2019 Feb 22;363(6429):810-812 [FREE Full text] [doi: 10.1126/science.aaw0029] [Medline: 30792287]
33. AlHogail A. Improving IoT technology adoption through improving consumer trust. Technologies 2018 Jul 07;6(3):64. [doi: 10.3390/technologies6030064]
34. Fridman I, Lucas N, Henke D, Zigler CK. Association between public knowledge about COVID-19, trust in information sources, and adherence to social distancing: cross-sectional survey. JMIR Public Health Surveill 2020 Sep 15;6(3):e22060 [FREE Full text] [doi: 10.2196/22060] [Medline: 32930670]
35. Oldeweme A, Märtins J, Westmattelmann D, Schewe G. The role of transparency, trust, and social influence on uncertainty reduction in times of pandemics: empirical study on the adoption of COVID-19 tracing apps. J Med Internet Res 2021 Feb 08;23(2):e25893 [FREE Full text] [doi: 10.2196/25893] [Medline: 33465036]
36. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
37. Lee TY, Bradlow ET. Automated marketing research using online customer reviews. J Mark Res 2011 Oct 01;48(5):881-894. [doi: 10.1509/jmkr.48.5.881]
38. Mankad S, Han HS, Goh J, Gavirneni S. Understanding online hotel reviews through automated text analysis. Serv Sci 2016 Jun;8(2):124-138. [doi: 10.1287/serv.2016.0126]
39. Day MY, Lee CC. Deep learning for financial sentiment analysis on finance news providers. 2016 Presented at: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 18-21, 2016; San Francisco, CA p. 1127-1134. [doi: 10.1109/asonam.2016.7752381]
40. Zhao W, Luo X, Qui T. Recent Developments in Smart Healthcare. Basel, Switzerland: MDPI; 2018.
41. Chen L, Wang P, Ma X, Wang X. Cancer communication and user engagement on Chinese social media: content analysis and topic modeling study. J Med Internet Res 2021 Nov 10;23(11):e26310 [FREE Full text] [doi: 10.2196/26310] [Medline: 34757320]
42. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
43. Liu L, Tang L, Dong W, Yao S, Zhou W. An overview of topic modeling and its current applications in bioinformatics. Springerplus 2016;5(1):1608 [FREE Full text] [doi: 10.1186/s40064-016-3252-8] [Medline: 27652181]
44. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2000 May 15;3(4-5):993-1022. [doi: 10.1162/jmlr.2003.3.4-5.993]
45. Bogoradnikova D, Makhnytkina O, Matveev A, Zakharova A, Akulov A. Multilingual sentiment analysis and toxicity detection for text messages. 2021 Presented at: 2021 29th Conference of Open Innovations Association (FRUCT); May 12-14, 2021; Tampere, Finland p. 55-64. [doi: 10.23919/fruct52173.2021.9435584]
46. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. 2012 Presented at: 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; 2012; Jeju Island, Korea.
47. GENSIM. Topic Coherence Pipeline. URL: https://radimrehurek.com/gensim/models/coherencemodel.html [accessed 2022-11-25]
48. Liu Q, Zheng Z, Chen J, Tsang W, Jin S, Zhang Y, et al. Health communication about hospice care in Chinese media: digital topic modeling study. JMIR Public Health Surveill 2021 Oct 21;7(10):e29375 [FREE Full text] [doi: 10.2196/29375] [Medline: 34673530]
49. Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Polit Anal 2017 Jan 04;21(3):267-297. [doi: 10.1093/pan/mps028]
50. Chuang J, Manning C, Heer J. Termite: visualization techniques for assessing textual topic models. 2012 Presented at: Proceedings of the International Working Conference on Advanced Visual Interfaces; May 2012; Capri, Italy p. 21-25 URL: https://dl.acm.org/doi/abs/10.1145/2254556.2254572 [doi: 10.1145/2254556.2254572]
51. Sievert C, Shirley K. LDAvis: a method for visualizing and interpreting topics. 2014 Presented at: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces; June 2014; Baltimore, MD p. 63-70. [doi: 10.3115/v1/w14-3110]
52. Ransing R, Nagendrappa S, Patil A, Shoib S, Sarkar D. Potential role of artificial intelligence to address the COVID-19 outbreak-related mental health issues in India. Psychiatry Res 2020 Aug;290:113176. [doi: 10.1016/j.psychres.2020.113176] [Medline: 32531628]
53. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005 Dec;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
54. Brown-Johnson CG, Boeckman LM, White AH, Burbank AD, Paulson S, Beebe LA. Trust in health information sources: survey analysis of variation by sociodemographic and tobacco use status in Oklahoma. JMIR Public Health Surveill 2018 Feb 12;4(1):e8 [FREE Full text] [doi: 10.2196/publichealth.6260] [Medline: 29434015]

55. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
56. Lee Y, Lin JL. The effects of trust in physician on self-efficacy, adherence and diabetes outcomes. Soc Sci Med 2009 Mar;68(6):1060-1068. [doi: 10.1016/j.socscimed.2008.12.033] [Medline: 19162386]
57. Harrison C. OK Google, Siri, Alexa, Cortana; Can you tell me some stats on voice search?.: Branded3 URL: https://edit.co.uk/blog/google-voice-search-stats-growth-trends/ [accessed 2022-11-25]
58. Seeger A, Heinzl A. Human versus machine: contingency factors of anthropomorphism as a trustinducing design strategy for conversational agents. In F Davis, R Riedl, J vom Brocke, P-M Leger, A Randolph (Eds. In: Davis F, Riedl R, vom Brocke J, Leger PM, Randolph A, editors. Information Systems and Neuroscience. Cham: Springer; 2018.
59. Angermeyer MC, Matschinger H. The stigma of mental illness: effects of labelling on public attitudes towards people with mental disorder. Acta Psychiatr Scand 2003 Oct;108(4):304-309. [doi: 10.1034/j.1600-0447.2003.00150.x] [Medline: 12956832]
60. Ran M, Zhang T, Wong IY, Yang X, Liu C, Liu B, CMHP Study Group. Internalized stigma in people with severe mental illness in rural China. Int J Soc Psychiatry 2018 Feb 28;64(1):9-16. [doi: 10.1177/0020764017743999] [Medline: 29183250]
61. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. BMJ 2007 Jul 07;335(7609):24-27 [FREE Full text] [doi: 10.1136/bmj.39246.581169.80] [Medline: 17615222]
62. Safran DG, Kosinski M, Tarlov AR, Rogers WH, Taira DA, Lieberman N, et al. The primary care assessment survey: tests of data quality and measurement performance. Med Care 1998 May;36(5):728-739. [doi: 10.1097/00005650-199805000-00012] [Medline: 9596063]
63. Pearson SD, Raeke LH. Patients' trust in physicians: many theories, few measures, and little data. J Gen Intern Med 2000 Jul;15(7):509-513 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.11002.x] [Medline: 10940139]
64. Hoffman RR, Johnson M, Bradshaw JM, Underbrink A. Trust in automation. IEEE Intell Syst 2013 Jan;28(1):84-88. [doi: 10.1109/mis.2013.24]
65. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]

Abbreviations

AI: artificial intelligence
DMHC: digital mental health care
HC: health care
LDA: latent Dirichlet allocation
MD: mental disorder
MH: mental health
MHC: mental health care
mHealth: mobile health
TM: topic modeling

Edited by A Kushniruk; submitted 16.04.22; peer-reviewed by J Huang, E Borycki; comments to author 28.06.22; revised version received 10.07.22; accepted 09.11.22; published 02.12.22.

Please cite as:
Shan Y, Ji M, Xie W, Lam KY, Chow CY
Public Trust in Artificial Intelligence Applications in Mental Health Care: Topic Modeling Analysis
JMIR Hum Factors 2022;9(4):e38799
URL: https://humanfactors.jmir.org/2022/4/e38799
doi:10.2196/38799
PMID:36459412

Original Paper

Answering Hospital Caregivers' Questions at Any Time: Proof-of-Concept Study of an Artificial Intelligence–Based Chatbot in a French Hospital

Thomas Daniel[1], MSc, PharmD; Alix de Chevigny[1], PharmD; Adeline Champrigaud[2], MSc; Julie Valette[2], MSc; Marine Sitbon[1], PharmD; Meryam Jardin[1], PharmD; Delphine Chevalier[1], PharmD; Sophie Renet[1,3], PharmD, PhD

[1]Department of Pharmacy, Paris Saint-Joseph Hospital Group, Paris, France

[2]Innovation and Transformation Department, Information Systems Directorate, Paris Saint-Joseph Hospital Group, Paris, France

[3]Learning, Training and Digital Education and Training Research Center, University of Paris Nanterre, Paris, France

Corresponding Author:
Sophie Renet, PharmD, PhD
Department of Pharmacy
Paris Saint-Joseph Hospital Group
185 Raymond Losserand Street
Paris, 75014
France
Phone: 33 144127191
Email: srenet@ghpsj.fr

Abstract

Background: Access to accurate information in health care is a key point for caregivers to avoid medication errors, especially with the reorganization of staff and drug circuits during health crises such as the COVID-19 pandemic. It is, therefore, the role of the hospital pharmacy to answer caregivers' questions. Some may require the expertise of a pharmacist, some should be answered by pharmacy technicians, but others are simple and redundant, and automated responses may be provided.

Objective: We aimed at developing and implementing a chatbot to answer questions from hospital caregivers about drugs and pharmacy organization 24 hours a day and to evaluate this tool.

Methods: The ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model was used by a multiprofessional team composed of 3 hospital pharmacists, 2 members of the Innovation and Transformation Department, and the IT service provider. Based on an analysis of the caregivers' needs about drugs and pharmacy organization, we designed and developed a chatbot. The tool was then evaluated before its implementation into the hospital intranet. Its relevance and conversations with testers were monitored via the IT provider's back office.

Results: Needs analysis with 5 hospital pharmacists and 33 caregivers from 5 health services allowed us to identify 7 themes about drugs and pharmacy organization (such as opening hours and specific prescriptions). After a year of chatbot design and development, the test version obtained good evaluation scores: its speed was rated 8.2 out of 10, usability 8.1 out of 10, and appearance 7.5 out of 10. Testers were generally satisfied (70%) and were hoping for the content to be enhanced.

Conclusions: The chatbot seems to be a relevant tool for hospital caregivers, helping them obtain reliable and verified information they need on drugs and pharmacy organization. In the context of significant mobility of nursing staff during the health crisis due to the COVID-19 pandemic, the chatbot could be a suitable tool for transmitting relevant information related to drug circuits or specific procedures. To our knowledge, this is the first time that such a tool has been designed for caregivers. Its development further continued by means of tests conducted with other users such as pharmacy technicians and via the integration of additional data before the implementation on the 2 hospital sites.

(JMIR Hum Factors 2022;9(4):e39102) doi:10.2196/39102

KEYWORDS

chatbot; artificial intelligence; pharmacy; hospital; health care; drugs; medication; information quality; health information; caregiver; healthcare staff; digital health tool; COVID-19; information technology

Introduction

Background

The hospital is a particularly complex environment that has to maintain continuous health care activities 7 days a week. A high staff rotation is essential to achieve permanent care, while ensuring proper clinical and pharmaceutical knowledge among staff members. This was even truer during the COVID-19 crisis, when many caregivers had to reorganize themselves and adapt to take care of patients. Moreover, the growing number of marketed drugs and their specificities (ie, drugs' availability, supply chain, dosage, compatibilities, stability, and administration techniques) has led to an increasingly large amount of available information that must be processed, far beyond practical use and knowledge for most caregivers and patients [1]. Besides, each hospital may have specificities in terms of processes, organization, and medication system. When a question about a drug arises, the health care staff, including nurses, needs to be able to find a full, reliable, and accurate answer to ensure proper and safe care. This research often has to be carried out in a short time in addition to the already existing workload, which further complicates the process [2]. Moreover, accurate information may not be easily accessible, especially during night shifts when there are fewer people available in the health care unit. Poorly documented websites might then end up being the last option to obtain information. This was all the more critical during the COVID-19 crisis when medical misinformation has abounded in numerous websites and forums, particularly about vaccines [3]. In such cases, the lack of access to accurate information can lead to medication errors [4,5]. Especially in hospital settings, any of these errors can have serious clinical consequences for patients.

The hospital pharmacy is then the main intermediary providing pharmaceutical advice. Nevertheless, telephone calls to the pharmacy can lead to a growing number of task interruptions for pharmacists, especially if these questions become repetitive, as a consequence of the high staff rotation, which is particularly frequent among nurses. Answers to these questions can also vary among pharmacists, depending on various factors such as the pharmacist's experience (senior, intern), his/her specialty, and the sources used for reference. Thus, there is a need for hospital pharmacists to find a new way to centralize, standardize, and internally verify the scientific validity of all answers to drug-related questions asked by caregivers. A computerized tool that would generate automated answers to caregivers, using the potential of artificial intelligence, could solve these issues.

Indeed, virtual assistants, also called chatbots, are already used in the field of human health. It is a very easy-to-use tool that users can access through SMS text messages, on smartphones, on computers, or on other connected devices (it needs an internet connection to function). It looks like a classic conversation window, and both the user and the chatbot can have a discussion in human language. The user asks his/her question directly in the dedicated field or clicks on specific buttons. Chatbots have been shown to be useful and effective to provide information and advice to patients with chronic conditions [6-8] as well as promoting a healthy lifestyle [9]. They are also used for oncological applications, in diagnosis and symptom screening, patient or treatment monitoring, as well as mental health counseling or emotional support [10-13]. However, there is currently no chatbot dedicated to caregivers' questions about drugs and pharmacy organization. We thought that a chatbot with a pharmaceutically validated database, implemented into the hospital intranet, could provide caregivers accessible and fact-checked answers in their everyday practice.

Objectives

We aimed at developing and implementing a chatbot, focusing on drugs and pharmacy organization to answer questions from hospital caregivers 24 hours a day and then evaluated it.

Methods

Setting

The proof-of-concept study took place in a large, private, nonprofit French hospital (comprising 592 beds) between 2021 and 2022 (1 year). The project was conducted by a multidisciplinary team composed of 3 hospital pharmacists, 2 members of the Innovation and Transformation Department, and an IT service provider.

Chatbot Design

Instructional Design

We used the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model [14] to set up the chatbot tool in the hospital—this is an instructional systems design model divided into 5 successive and iterative phases.

Analysis

During 1 month in February 2021, we asked hospital caregivers from 5 departments—especially nurses and pharmacists—what the most frequently asked questions about drugs and pharmacy organization were in their daily practice. We then synthetized the results to integrate appropriate items into the chatbot.

Design and Development

We designed and introduced relevant topics into the chatbot in accordance with the previous needs analysis. To provide reliable and clear answers, different sources of information were cross-checked. The following are various pharmaceutical professional tools at our disposal in France: pharmaceutical references derived from the summaries of product characteristics; local recommendations of the Commission for Medicinal Products and Sterile Medical Devices of the establishment; publications of health authorities such as the Observatories of Medicines, Medical Devices and Therapeutic Innovations, the French National Agency for Drugs and Medicinal Products Safety, and the French National Authority for Health; and reference websites such as Stabilis. Each working document was reviewed and scientifically verified by a hospital pharmacist in the relevant field and then implemented into the chatbot. For drug names, we integrated both international nonproprietary names and brand names in the chatbot, so the tool understood both formulations.

We then developed the chatbot itself, using the node.js runtime environment and typescript language, through regular exchanges with the IT service provider. In order for the chatbot to understand human language and be able to answer, we also used Watson's application programming interface (IBM Corp). Thus, the chatbot used natural language processing and machine learning to understand the user and learn from their previous interactions. It also used an auto-completion engine developed with the IT provider to help users write the drugs' names. Hence, when the user entered the first letters of a drug's name, a list of matching proposals was displayed. Moreover, we integrated a fuzzy matching feature to avoid spelling differences and mistakes.

Implementation and Evaluation

We implemented the chatbot in the hospital intranet of 4 tester departments to evaluate it. Hence, the beta test was conducted between January and February 2022 before implementing the tool in all hospital care units. It consisted of 20-minute trial sessions organized within the hospital care services, at the end of which, testers (head nurses, nurses, pharmacy assistants, pharmacy students, and interns who agreed to participate) were asked to complete a questionnaire about themselves and about their experience with the chatbot. They had to rate the chatbot's speed, usability, and appearance, and they were asked to assess their satisfaction with it on a 5-point Likert scale (from "very unsatisfied" to "very satisfied"). They were also asked what missing topics they wanted to include in the chatbot to improve the tool. In addition, we also measured the performance of the tool from the IT provider's back office. At the end of each interaction with the chatbot, users were invited to provide positive or negative feedback (thumb up or down). The relevance of the chatbot was defined as the number of positive feedbacks compared to the number of total expressed feedbacks, and we also counted the number of nonexpressed feedbacks.

Ethics Approval

No ethics approval was necessary for this study, given that the chatbot does not fall within the definition of a medical device within the meaning of European Union regulation 2017/745, and that French legislation (public health code) does not require ethics approval for this kind of work.

Results

Needs Analysis

The first step, which consisted of collecting and analyzing the needs of future users, was conducted in the hospital in February 2021. In total, 50 questions asked by 33 nurses, 3 head nurses, and assistant nurses were collected and classified into 7 categories. Five different health care units were included: Oncology Outpatient Service, Oncology Inpatient Service, Weekday Inpatient Oncology Service (WIOS), Pneumo-Oncology Service, and Neonatology Service. We also took into consideration the suggestions of 5 hospitals pharmacists concerning the drug circuits and anticancer drugs. Injectable drug compatibility, stability, and administration methods were the main subjects of these questions. Other topics included therapeutic equivalences and anticancer drugs, as shown in Table 1.

Table 1. Topics of the caregivers' questions during the needs analysis.

	Oncology Outpatient Service	Oncology Inpatient Service	Weekday Inpatient Oncology Service	Pneumo-Oncology Service	Neonatology Service	Total, n (%)
Administration, n	—[a]	3	7	3	—	13 (26)
Availability, n	—	1	—	1	—	2 (4)
Anticancer drugs, n	4	—	—	—	—	4 (8)
Compatibility and stability, n	3	4	1	5	1	14 (28)
Cutting or crushing tablets, n	—	2	—	3	—	5 (10)
Miscellaneous, n	5	1	—	2	2	10 (20)
Therapeutic equivalences, n	—	1	1	—	—	2 (4)
Total, n (%)	12 (24)	12 (24)	9 (18)	14 (18)	3 (6)	50 (100)

[a]Not determined.

Design and Development

The 50 topics indicated during the needs analysis led to the writing of 32 working documents, gathered by item. From the working documents, we created 41 skills into the chatbot, which formed the database with the corresponding resource documents. For each skill, we also built a conversational tree, which constituted a decisional algorithm to determine how the chatbot might interact with the user to provide the requested response (Figure 1). For example, depending on the type of question asked, the chatbot might ask the user for further information, answer directly, or provide a web link to find the correct answer on a reliable website. An example of how the chatbot handles a request concerning the availability of medication is detailed in Multimedia Appendix 1. The chatbot was designed to function in French. We did not provide the tool a name, as it was intended to be integrated later into a global chatbot (meta-bot); we simply called it "Pharmacy chatbot."

We also included in the chatbot an initial dialogue to guide the users more quickly on frequently asked topics and used technical solutions called "fuzzy matching" and "auto completion" to help users formulate their requests and make them understandable by the chatbot (Figure 2).

Figure 1. Pharmacy chatbot design scheme. Construction of the chatbot was based on the analysis of the needs of caregivers to integrate relevant data. To understand the questions asked, a natural language processing application programming interface (Watson API, IBM Corp) and a homemade dictionary were also included. API: application programming interface.

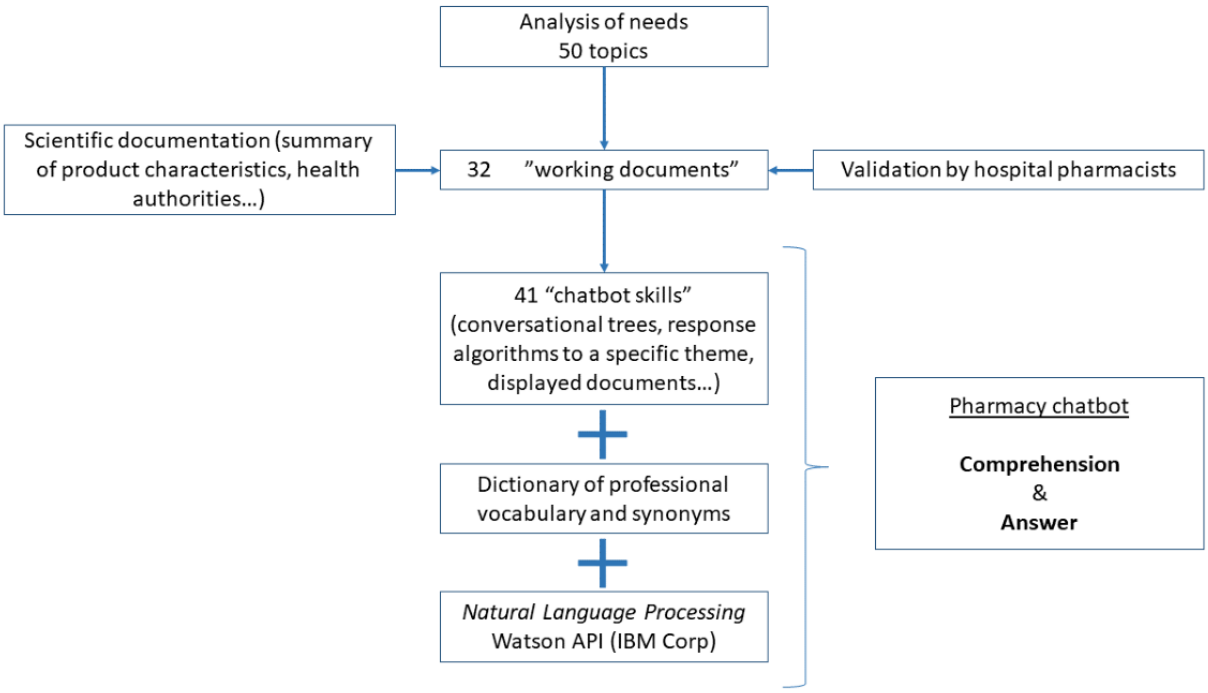
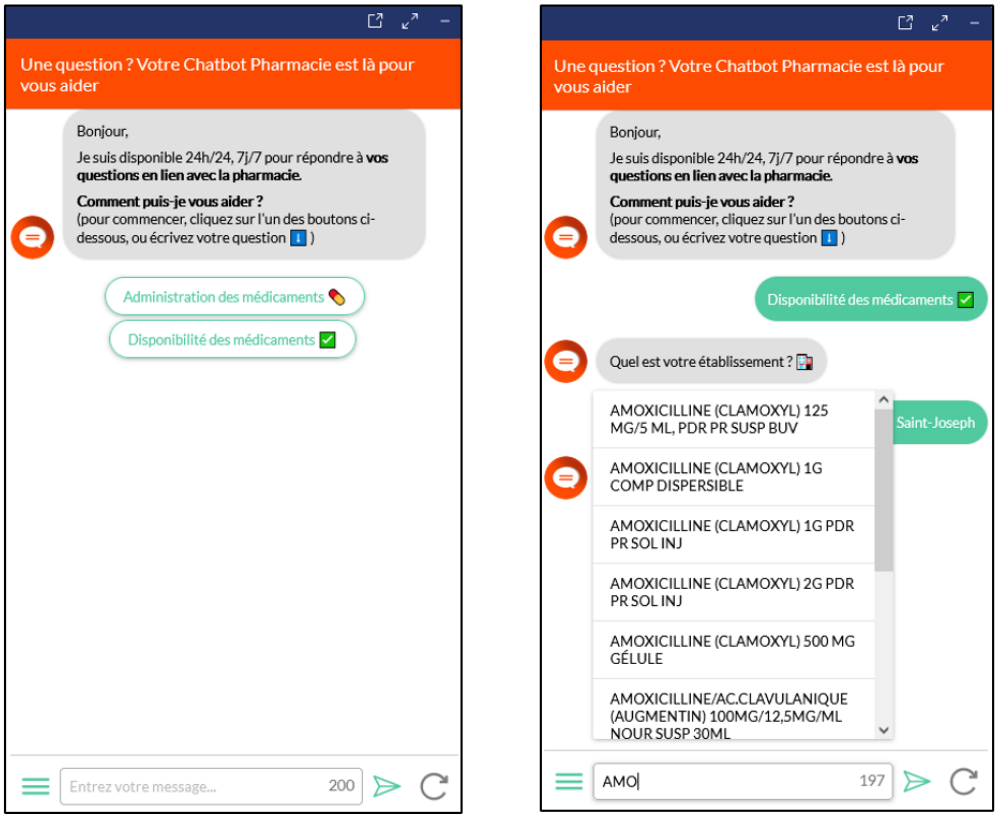


Figure 2. Screenshots of the pharmacy chatbot interface. Left: the opening dialogue is designed to guide the users on frequently discussed topics. Right: the auto completion system helps users write drug names correctly.



Implementation and Evaluation

The beta test of the chatbot tool was conducted between January and February 2022. In total, 20 caregivers from 4 different services (Oncology Outpatient Service, WIOS, Pneumo-Oncology Service, and pharmacy department) attempted the proof-of-concept version of the chatbot during test sessions. A total of 14 nurses and head nurses participated, as well as 6 members of the pharmacy staff. The beta test led

to 214 conversations, and testers were invited to complete a satisfaction questionnaire, the results of which are presented in Figure 3.

Overall, 8 of 20 (40%) testers used to call the pharmacy 1 to 5 times a week, and 11 (55%) used to do it more. Only one person, an experienced head nurse, said that she never had to call the pharmacy. The chatbot's speed was rated 8.2 out of 10 (range 3-10), its ergonomics was rated 8.1 (range 5-10), and its appearance was rated 7.5 (range 4-10) as shown in Figure 4. One person did not rate the chatbot. Overall, 14 of 20 (70%) users were satisfied or very satisfied with the tool. In the back office, the estimated relevance (ie, conversations with positive feedback) was 76% for of interactions that were not aborted (146/214, 68% of conversations).

At the end of the beta tests, the main improvements that were suggested by the testers were related to the ergonomics (n=5, 25% of testers), as well as the implementation of the database (n=7, 35% of testers), with the inclusion of answers related to medical devices (n=4, 20% of testers), adverse drug effects (n=4, 20% of testers), and drug prices (n=2, 10% of testers), as shown in Figure 5.

Figure 3. Descriptive analysis of the population of beta testers. Testers were invited to respond a questionnaire with their attitudes toward the previous use of a chatbot, their estimated familiarity with the computer tool on 5-point Likert scale, and the weekly number of calls to the hospital pharmacy. Testers were also invited to indicate their care unit and their experience.

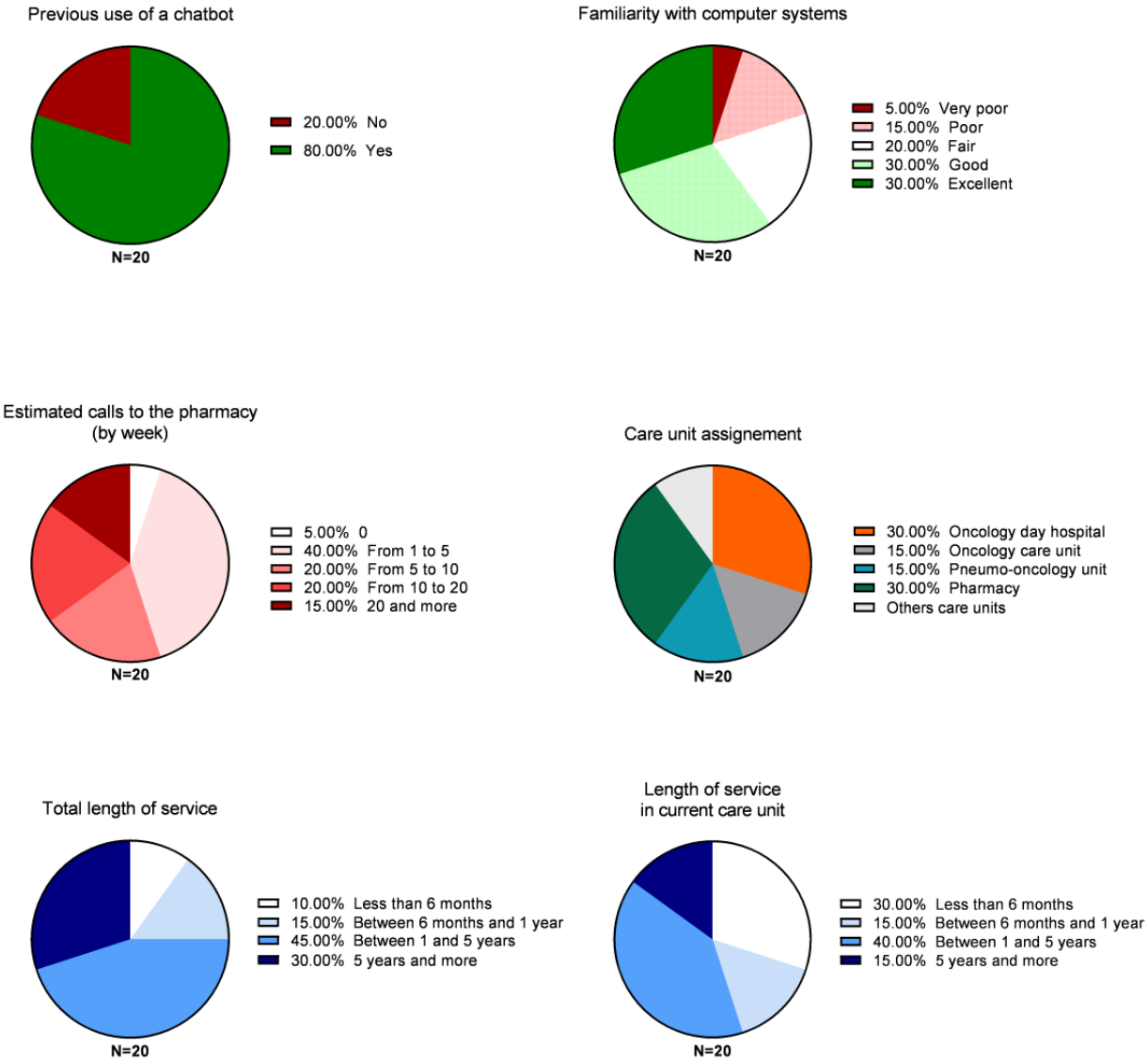


Figure 4. Evaluation scores of the chatbot appearance, speed, and usability after the beta test. Each point matches with an individual value, and vertical lines represent the mean and SD values.

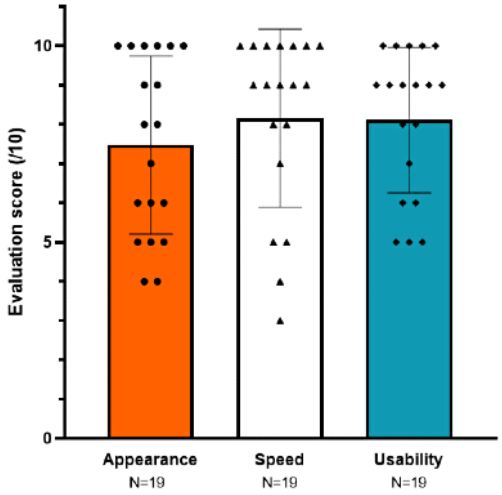
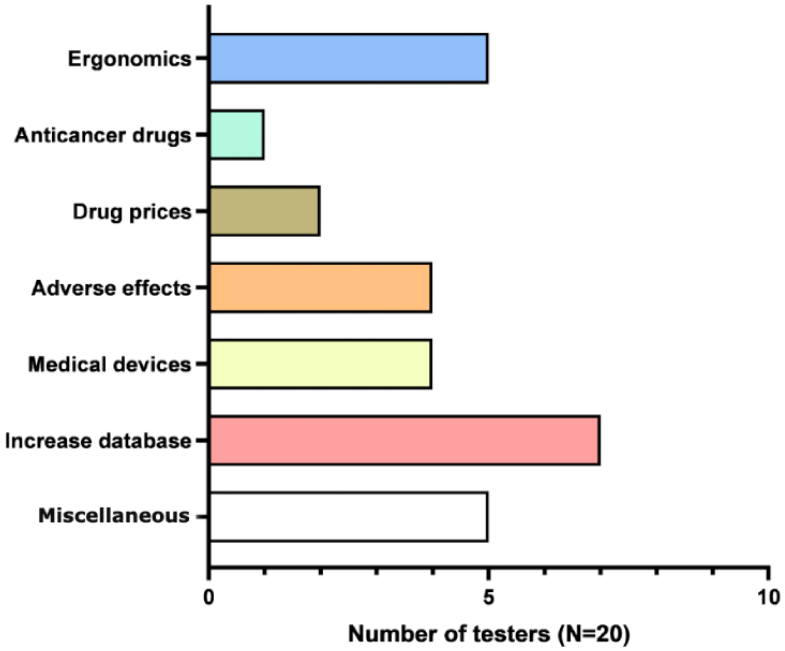


Figure 5. Suggestions for future improvements (by topic) from testers after the trial session. Percentages were calculated on the basis of the 20 people who took part.



Discussion

Principal Findings

To our knowledge, the chatbot we designed is the first to specifically target the hospital health staff and answer questions related to drug circuits and pharmacy organization. There are currently very few chatbot-type solutions specifically dedicated to the hospital setting, and almost all of the existing literature deals with chatbots in relation to patients [15,16]. The perception of the chatbot tool by physicians has already been studied [17]. In this case, the chatbot made it possible to provide prescription aid but also information on drugs, such as our chatbot. In particular, the authors concluded that the chatbot would perfectly fit in the daily medical practice and that it was positively perceived by the physicians. However, the main limitation of the work remained the lack of information included in the database.

During the trial session, the beta testers found our chatbot to be fast, user-friendly, and aesthetically pleasing. The overall satisfaction as well as the actual performance of the chatbot were considered good (all scores were higher than 7 out of 10), and all testers explained that they would want to continue using the tool if its database is widened. They also declared that the chatbot could help them during the COVID-19 crisis to obtain reliable information they need about vaccination and easily respond to patient requests. The database was thus improved to take into account the suggestions of testers, and in doing so, expand the area of competence of the chatbot. We, therefore, added topics regarding drug prices, adverse drugs effects, and some questions dealing with anticancer drugs. We also added recent data regarding the management of vaccination of pregnant women, in accordance with the French Reference Center on Teratogens recommendations.

Thus, the chatbot solution developed and implemented in our hospital appears to be a useful and reliable tool to address the

common drug- and pharmacy-related questions encountered by nurses and other hospital caregivers at any time during their practice. It is accessible, implemented in the hospital intranet that health professionals are accustomed to using, and easy to use.

Limitations

We attempted to include as much data as possible in the chatbot system to answer caregivers' questions, taking into account the limited resources (both human and technical) allocated to its development. The small number of beta testers (n=20) is, therefore, a limitation of our work, and there are also other data that need to be integrated into the chatbot. One of the most frequent requests concerned drugs' stability. To provide an answer, we included cross-tables that are concise, extremely dense, and not very user-friendly. Another solution would be to redirect the user to a reliable website such as specialized websites, but this would force users to transcribe their information and use another tool, which is highly likely to lead to a loss of support toward the chatbot. Finally, the topic of medical devices is also under study; however, the main difficulty remains the complex nomenclature and the extremely variable names that a single device may have. This topic has not been included in the chatbot's scope, and currently, the chatbot simply invites users asking a question about it to contact specifically pharmacists or pharmacy assistants working in this field.

Another technical limitation of the chatbot tool is the fact that it was not possible to interface the chatbot with our business software because of an incompatible application programming interface. This limitation is likely to evolve in the future, with software companies taking into account the need for flows' interoperability.

When interpreting the results, it is also important to note that the relevance score evaluated during beta tests might be overestimated because of interactions that were interrupted by testers before the end of the study period. These aborted interactions are not recorded in the performance ratio included in the back office.

On the user's side, the degree of mastery of computer tools generally determines the handling and the acceptance of the chatbot. There is also resistance to change, as the use of the chatbot is not currently part of the habits and routine of the staff. It is, therefore, necessary to facilitate its use as much as possible, and its relevance should be demonstrated in the early stages of use to promote its integration into the caregivers' practice. Caregivers will include the chatbot in their practice in the long run only if its effectiveness and time-saving potential during searches for reliable information are clearly demonstrated. Should it fail to demonstrate these abilities, the deceptive effect is likely to prevail, and beyond the "novelty effect," caregivers may lose interest in the project.

Perspectives

Regular updating of the information integrated in the database is a central issue in maintaining a high level of reliability. This requires the appointment of a person who would be directly responsible for managing the chatbot. Thus, it is possible to entrust this task to a pharmacy intern, who could weekly consult the unanswered questions and integrate the corresponding information into the chatbot. We expect to implement these updates on a monthly to biannual basis according to the frequency of modification of the information and their levels of criticality. Moreover, the chatbot manager could also inform users in real time about significant breaks or important news, using the "programmed events" feature. It consists of a banner—that would appear when the chatbot is launched—to feature important information. However, this would require close collaboration with the procurement sector of the hospital pharmacy.

In the future, it may also be possible to expand the scope of the chatbot's skills even further, by including, for example, a list of medical devices available in the hospital. However, given the extremely large number of references, it will be necessary to implement a relatively effective indexing or search system to better guide the user among the complex diversity of labels. Another major element in the future of the chatbot is its deployment on the 2 hospital sites so that all hospital caregivers can use it. In order to do so, it is necessary to take into account the specificities of each site; for example, drug availability. In another update, we could also include a list of verified, up-to-date websites to convert this tool to a weapon against medical misinformation. In the context of health crises, the chatbot may be a relevant means of information to transmit homogeneous information to all caregivers of our hospital. The development of different type of interfaces, such as smartphones or a speech-to-text option, may also be a relevant perspective to improve the ergonomics and adherence of caregivers.

Finally, the chatbot could become an effective training tool for pharmacy technicians. Indeed, the chatbot would allow them to answer a wide variety of drugs-related questions on their own and more easily, without necessarily having to ask a referent pharmacist. Of course, questions about the patient's clinic or specific elements of drugs will always require pharmaceutical expertise, but we believe that the chatbot may allow pharmacists to optimize their time for value-added tasks. The tool could also constitute an essential element for the training of new staff and for personnel who may move between care units for organizational purposes.

Conclusions

The chatbot seems to be a useful tool for hospital caregivers, helping them obtain the information they need about drugs and pharmacy organization. To our knowledge, this is the first time that such a tool has been designed for hospital caregivers. It is, therefore, crucial to continue developing it using tests with and implementing additional data to improve its performance. Several aspects still need to be improved, especially in relation to the scope of the chatbot's skills. Eventually, the technology could be shared with other institutions, or even to websites, such as those of learned societies. The chatbot could also become an essential component for the training of new health care staff and on the occasion of staff changes in different departments for organizational purposes.

Acknowledgments

We deeply thank the members of the IT service provider (especially Marion Schaeffer, Christophe Bouvard, and Simon Roch) who helped us develop the chatbot. We also thank others who reviewed and helped proofread the manuscript (especially MS, DC and Sophie Daniel), as well as all people who tested the chatbot.

Authors' Contributions

TD and SR contributed to the study’s conception and design, as well as material preparation, data collection, and analysis. Beta tests were organized and conducted by TD, SR, AdC, and JV. TD wrote the first draft of the manuscript, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Schema of the chatbot's decision algorithm regarding the availability of a drug.
[PNG File , 73 KB - humanfactors_v9i4e39102_app1.png]

References

1. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? Z Evid Fortbild Qual Gesundhwes 2015;109(4-5):285-290. [doi: 10.1016/j.zefq.2015.06.005] [Medline: 26354128]
2. Verhoeven F, Steehouder MF, Hendrix RMG, van Gemert-Pijnen JEWC. How nurses seek and evaluate clinical guidelines on the internet. J Adv Nurs 2010 Jan;66(1):114-127. [doi: 10.1111/j.1365-2648.2009.05175.x] [Medline: 20423439]
3. Bin Naeem S, Kamel Boulos MN. COVID-19 misinformation online and health literacy: a brief overview. Int J Environ Res Public Health 2021 Jul 30;18(15) [FREE Full text] [doi: 10.3390/ijerph18158091] [Medline: 34360384]
4. Hammoudi BM, Ismaile S, Abu Yahya O. Factors associated with medication administration errors and why nurses fail to report them. Scand J Caring Sci 2018 Sep;32(3):1038-1046. [doi: 10.1111/scs.12546] [Medline: 29168211]
5. Nurmeksela A, Mikkonen S, Kinnunen J, Kvist T. Relationships between nurse managers' work activities, nurses' job satisfaction, patient satisfaction, and medication errors at the unit level: a correlational study. BMC Health Serv Res 2021 Apr 01;21(1):296 [FREE Full text] [doi: 10.1186/s12913-021-06288-5] [Medline: 33794875]
6. Echeazarra L, Pereira J, Saracho R. TensioBot: a chatbot assistant for self-managed in-house blood pressure checking. J Med Syst 2021 Mar 15;45(4):54. [doi: 10.1007/s10916-021-01730-x] [Medline: 33723721]
7. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
8. Kökciyan N, Chapman M, Balatsoukas P, Sassoon I, Essers K, Ashworth M, et al. A collaborative decision support tool for managing chronic conditions. Stud Health Technol Inform 2019 Aug 21;264:644-648. [doi: 10.3233/SHTI190302] [Medline: 31438003]
9. Oh YJ, Zhang J, Fang M, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021 Dec 11;18(1):160 [FREE Full text] [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]
10. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850 [FREE Full text] [doi: 10.2196/27850] [Medline: 34847056]
11. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019 Nov 27;21(11):e15787. [doi: 10.2196/15787] [Medline: 31774408]
12. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
13. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
14. Ingénierie pédagogique et modèle ADDIE. Sydologie. 2017 Sep 28. URL: http://sydologie.com/2017/09/ingenierie-pedagogique-modele-addie/ [accessed 2021-05-19]
15. Gabarron E, Larbi D, Denecke K, Årsand E. What do we know about the use of chatbots for public health? Stud Health Technol Inform 2020 Jun 16;270:796-800. [doi: 10.3233/SHTI200270] [Medline: 32570492]

16. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
17. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]

Abbreviations

ADDIE: Analysis, Design, Development, Implementation, and Evaluation
WIOS: Weekday Inpatient Oncology Service

Edited by A Kushniruk; submitted 28.04.22; peer-reviewed by A Palanica, P Nitiema; comments to author 18.07.22; revised version received 24.07.22; accepted 02.08.22; published 11.10.22.

Please cite as:
Daniel T, de Chevigny A, Champrigaud A, Valette J, Sitbon M, Jardin M, Chevalier D, Renet S
Answering Hospital Caregivers' Questions at Any Time: Proof-of-Concept Study of an Artificial Intelligence–Based Chatbot in a French Hospital
JMIR Hum Factors 2022;9(4):e39102
URL: https://humanfactors.jmir.org/2022/4/e39102
doi:10.2196/39102
PMID:35930555

Original Paper

Food Delivery Drivers' Health Literacy Regarding COVID-19 Prevention and Protective Behaviors During the COVID-19 Pandemic: Cross-sectional Survey in Southern Thailand

Kasemsak Jandee[1,2], PhD; Chamnong Thanapop[1,2], PhD

[1]Department of Community Public Health, School of Public Health, Walailak University, Nakhon Sri Thammarat, Thailand

[2]Center of Excellence in Data Science for Health Study, Walailak University, Nakhon Sri Thammarat, Thailand

Corresponding Author:
Kasemsak Jandee, PhD
Department of Community Public Health
School of Public Health
Walailak University
222 Thaiburi
Thasala District
Nakhon Sri Thammarat, 80160
Thailand
Phone: 66 7 567 6671
Fax: 66 7 567 2705
Email: kasemsak.ja@wu.ac.th

Abstract

Background: In 2019, COVID-19 spread worldwide, causing a pandemic that has posed unprecedented challenges and pressure for health systems and economies. Food delivery services have become an important medium for consumer food purchases to limit human-to-human contact. Thus, delivery drivers are at high risk of exposure to COVID-19 infection at work. To the best of our knowledge, no studies have analyzed the dimensions of health literacy (HL) regarding COVID-19 prevention in this population.

Objective: This study aims to explore the HL status toward COVID-19 prevention and its associated factors among food delivery drivers in southern Thailand.

Methods: Following a cross-sectional survey from July to August 2021, Thai food delivery drivers in the upper-south and lower-south regions of southern Thailand were recruited to participate during the compulsory COVID-19 lockdown. An online structured questionnaire was administered verbally and recorded by the interviewer. Univariate and multivariate linear regressions were used to explore independently associated factors.

Results: Of 401 drivers, 291 (72.6%) were men. The median age was 31 years (range 19-64 years). The median number of months working as a driver was 12 months, and the median number of working hours was 9 hours per day. The median number of daily food orders was 20, while the median daily income was Thai baht (THB) 600 (US $15.90). Social media (Facebook and Line) was a common source of health information. The most common information required was about the COVID-19 vaccine, medications, and treatment. Most drivers (285/401, 71.1%) had excellent HL levels regarding COVID-19 prevention. Only the practical application of information was statistically correlated with behavior (r=0.38, P<.001). Drivers in the lower south of Thailand were more likely to have excellent HL than other drivers (β=7.03, P<.001). Those who frequently accessed information through YouTube (β=–2.17, P=.01) and relatives (β=–4.19, P<.001) were less likely to have excellent HL levels.

Conclusions: Understanding HL among food delivery drivers would be useful for planning effective interventions that target this population. Conventional health education through social media alone may not be effective at educating people about COVID-19 prevention. Information literacy skills could determine individuals' HL and drivers' behaviors.

(JMIR Hum Factors 2022;9(4):e37693) doi:10.2196/37693

KEYWORDS

health literacy; health information; preventive behavior; COVID-19; food delivery; delivery drivers

Introduction

Since the onset of the COVID-19 pandemic that has been declared a global emergency by the World Health Organization (WHO) [1,2], social distancing and self-isolation, including lockdowns, have been adopted almost universally by countries worldwide as public health and social measures to control the transmission of the virus [3]. These changes have occurred through citizens' voluntary steps and concrete government interventions. With the global recommendation of maintaining social distancing and implementing lockdowns worldwide, many businesses, especially restaurants, have been forced to close, particularly in Thailand. Thus, food delivery services have become an important medium for consumer food purchases [4,5], and the pandemic has significantly changed the food delivery service industry and consumer perceptions [6].

People generally perceive that using food delivery services is safer than going into restaurants because the limited human-to-human contact reduces the risk of SARS-CoV-2 infection [7]. A previous study revealed that consumers are less concerned about contracting COVID-19 from food in general than they are for restaurant food [8]. However, delivery drivers are a highly mobile population that offers services to a wide range of clients, including vulnerable populations, such as older adults or those less likely to leave their house for stocking up on basic needs [9]. In Thailand, there was no formal guideline or operating procedures for the drivers in the context of COVID-19 mentioned by the government or food delivery companies, and the strictness of social and public health measures was different depending on the rapid transmission of COVID-19 and cluster infections in each province.

Food delivery services in Thailand have been estimated to be worth over US $1.1 billion, with a 17% growth rate for 2020. The 4 major food delivery service providers in Thailand are GrabFood, Food Panda, Line Man, and Get [10]. Thai food delivery has been growing gradually at approximately 10% since 2017, and the COVID-19 pandemic has catalyzed growth by limiting Thai people’s ability to eat at restaurants. The pandemic has benefited Thai food delivery service operators as orders from customers have increased by 100%-300%, with at least 225,000 delivery drivers working in Thailand [11]. Drivers have suddenly been thrust into the front line of the COVID-19 pandemic. Due to the increased movement of food delivery drivers and the high number of people with whom they come into contact, these drivers are at risk of exposure to SARS-CoV-2 infection at work, similar to health care providers. Furthermore, food delivery drivers who are suspected of having contracted SARS-CoV-2 may play a role in actively transmitting the virus to consumers [12,13].

The number of COVID-19 cases among food delivery drivers has been possibly underreported by food delivery companies. A cross-sectional study conducted in Quito, Ecuador, in 2020 revealed a high incidence rate of SARS-CoV-2 infection in self-employed food delivery drivers [9]. A food delivery man in Beijing, China, was reported as having SARS-CoV-2 and had a record of delivering around 50 orders per day across a wide area in Beijing in June 2020 [14]. In addition, a previous report from a public hospital in Hanoi, Vietnam, showed that more than 60% of confirmed COVID-19 cases could be linked to food delivery at the hospital’s cafeteria [15]. Another case was reported in India, where a pizza delivery man tested positive for SARS-CoV-2. His history showed that he had contact with 72 families, including 17 other delivery men, all of whom were immediately quarantined [16]. The lack of appropriate occupational health control measures for food delivery drivers and their movements and social interactions, which are high-risk behaviors, put this population at high risk of contracting COVID-19. However, food delivery drivers are expected to play a significant role in reducing the risk of COVID-19 transmission [17].

To ensure that people adhere to infection control precautions during the COVID-19 pandemic, they must be able to access and understand public health information. People’s ability to obtain, use, and apply information to make decisions related to their health is defined as health literacy (HL) [18,19]. Many researchers have found that better public health outcomes result from people’s acquisition of new knowledge, and more positive attitudes, greater self-efficacy, and positive health behaviors are associated with higher HL [18].

Regarding HL tools, a previous systematic review showed an upward trend in the availability of tools assessing HL among the general population using multidimensional structures and comprehensive measurement approaches. However, a definite consensus could not be reached on the dimensions of HL tools [20]. Thus, an appropriate HL tool should be developed according to the specific health condition being addressed. In addition, understanding the drivers’ information needs and information-seeking behaviors is essential for developing information systems and services that adequately satisfy their needs. However, to the beset of our knowledge, there have recently been no studies exploring the dimensions of food delivery drivers’ HL toward COVID-19 prevention, even though food delivery services have been growing annually during the COVID-19 pandemic, particularly in Thailand. To investigate how food delivery drivers in southern Thailand recognize COVID-19 prevention, this study was conducted to explore their HL status toward COVID-19 prevention and its association with sociodemographic characteristics, work-related factors, and health information access. The findings from this study could provide scientific evidence for COVID-19 control and prevention programs, including suggestions for public health information campaigns for HL for food delivery services during the COVID-19 pandemic.

Methods

Study Design and Setting

We conducted a cross-sectional survey between July and August 2021 in southern Thailand, a long, narrow peninsula that can be further divided into an upper-south and a lower-south region [21]. Two provinces, Nakhon Sri Thammarat and Songkhla, were selected as representative of the upper-south and lower-south regions, respectively, due to the high incidence of SARS-CoV-2 infection in these provinces during the duration of this study [22]. These provinces were also selected due to

the growth of food delivery services expanding from Bangkok [11]. This study was conducted using the KoBo Toolbox, a free and open source software for online and offline data collection developed by the Harvard Humanitarian Initiative [23].

Study Participants

Thai food delivery drivers aged 18 years or older who were food delivery drivers for at least 3 months before the administration of the survey were invited to participate in the study. Those who could not read or understand Thai and those who had a communication barrier, such as being deaf or having a mental deficit, were excluded.

The required sample size was determined using the single proportion formula according to the estimated proportion of adequate HL among food delivery drivers at 50% [24] since there were no previous studies conducted in this study population in Thailand, with a 5.5% acceptable error rate and a 95% CI. Since this cross-sectional survey used convenience sampling because food delivery drivers' interview time is a constraint, the design effect was not considered. An additional 20% of subjects were included to prevent data loss. Thus, the sample size required was at least 382 participants. The study met this requirement and included 401 participants (401/417, 96.2% response rate). Through intensive outreach with the snowball sampling technique in each region, the recruitment process was initiated by spreading information about the study through a group of food delivery drivers. The first interested individual was invited to participate in this study and helped us identify further potential participants. This step was repeated until the needed sample size was found.

Instrument and Measurements

A structured questionnaire was developed to assess the HL status of Thai food delivery drivers regarding COVID-19 prevention. The questionnaire consisted of sets of questions to determine participants' demographics, work-related factors, health information access, understanding, judgment, and application of health information toward COVID-19 prevention, including preventive behaviors. The questionnaire was reviewed by 3 public health experts, who rated the overall content validity of the questionnaire at 1.00, where the index of item objective congruence was over 0.5. A pilot test of the survey instruments was then conducted among 30 participants prior to administering the survey. The pilot study returned acceptable reliability values for each dimension of HL (0.73-0.81). The pilot study survey was verbally administered in the Nakhon Sri Thammarat province in southern Thailand, and the interviewer recorded responses on electronic handheld devices using KoBo technology [23]. This device can be used for data collection offline and then synchronized onto a central database when telephone signals or wireless networks are available. It should be noted that the personal information was treated confidentially within the system applications during the study period.

Questions regarding demographic characteristics included age, sex, education, food delivery company, religion, income, years/months of experience in food delivery services, number of hours worked per day, and number of completed food orders delivered per day. A single letter was used to represent each food delivery company's name to avoid disclosure of the official name due to this being a sensitive issue resulting in a competitive advantage. Questions on current health status included pre-existing conditions, cigarette smoking, and alcohol consumption. Participants' sources of health information were requested to determine their recent information access about COVID-19 prevention. The 12 questions regarding understanding were yes/no questions, where correct answers were given a score of 1, while incorrect answers were given a score of 0. Therefore, the participants' understanding scores ranged from 0 to 12. A set of questions to measure how the participants make decisions when confronted with COVID-19 prevention information was used to indicate the degree of participants' ability to judge/make decisions on each statement. Participants were asked to rate 6 questions on a 5-point Likert scale, with the responses being "extremely easy" (5), "slightly easy" (4), "neutral" (3), "slightly difficult" (2), and "extremely difficult" (1). Participants' ability to apply health information to their profession was assessed using 5 questions. Items were answered using a 5-point Likert scale from 1 (never) to 5 (always); total scores ranged from 5 to 25. The frequency of certain COVID-19 prevention behaviors was assessed using a 5-point Likert scale: "never" (1), "seldom" (2), "sometimes" (3), "often" (4), and "always" (5). COVID-19 prevention behavior scores ranged from 6 to 30. The total score for each section was summarized, and HL levels were classified according to the criteria. As there is no specific guideline for how to classify HL levels [20,25], the overall HL levels and other dimensions, including the levels of preventive behaviors, were classified as excellent (score≥80%), moderate (score=60%-79%), and inadequate (score<60%).

Statistical Analysis

Demographic, work-related, and health information–seeking factors of food delivery drivers were descriptively presented as percentages, mean (SD), or median (IQR). HL was determined using the sum of the scores from the understanding, judgment/decision-making, and application of information questions. Access to COVID-19 prevention information was described in terms of health information sources commonly used by the participants. This information was not included in the overall HL score.

Associations between the HL score, preventive behavior score, and independent variables were assessed using univariate/multivariate linear regression. A stepwise multiple linear regression model was used to determine the most significant predictors of HL and behavior toward COVID-19 prevention. We described the strength of the measure of association using the mean difference in the regression analysis. Correlation analyses between the components of HL and preventive behaviors among drivers were performed using the Pearson correlation coefficient (r). Differences were considered statistically significant at a *P* value of .05. All statistical analyses were performed using R version 4.1.2 statistical analysis software (R Foundation for Statistical Computing).

Ethical Considerations

This study was reviewed and approved by the Human Research Ethics Committee of Walailak University, Thailand, and

followed the principles of the Declaration of Helsinki (WUEC-21-071-01). All participants were informed of all details regarding the study, and informed consent was obtained before the participants completed the online form.

Results

Demographic and Work-Related Factors

A total of 401 participants completed an online questionnaire survey, and nearly three-quarters of them (291/401, 72.6%) were male. Demographic characteristics and work-related factors of the participants are presented in Table 1. The participants' age ranged from 19 to 64 years, with a median age of 31 years. Of the 401 participants, 344 (85.8%) were Buddhist, and most of the participants' education level was high school or lower. Company A represents the highest category of food delivery company. Participants reported that they were food delivery drivers for a median of 12 months. The median number of working hours was 9 hours per day, the median number of food order deliveries per day was 20, and participants' median daily income was Thai baht (THB) 600 (US $15.90). Most participants' (241/401, 60.1%) health insurance status was a universal health coverage scheme. More than half of the participants did not smoke (245/401, 61.1%) or consume alcohol (272/401, 67.8%). In addition, 21 (5.2%) of the participants had pre-existing chronic illnesses. The distributions of these characteristics among participants in the upper-south (n=201, 50.1%) and lower-south (n=200, 49.9%) regions of Thailand were almost all different, excluding pre-existing chronic illness status and alcohol consumption, as shown in Table 1. Regarding work-related factors, participants in the lower south of Thailand were more likely to have a longer working duration than those in the upper-south region (12 vs 8 months, $P<.001$). In contrast, participants in the upper south of Thailand were more likely to deliver a higher number of food orders (20 vs 18 deliveries, $P<.001$) and earn daily income (THB 600 vs 500 [US $15.90 vs $13.25], $P<.001$) than those in lower-southern Thailand.

Table 1. Demographic characteristics and work-related factors of participants.

Characteristics	Total (N=401)	Upper south (n=201)	Lower south (n=200)
Sex, n (%), *P*[a]=.002			
Male	291 (72.6)	160 (79.6)	131 (65.5)
Female	110 (27.4)	41 (20.4)	69 (34.5)
Age (years), *P*<.001			
Median (IQR)	31 (24-40)	27 (23-34)	36 (28-46)
Religion, n (%), *P*<.001			
Buddhism	344 (85.8)	186 (92.5)	158 (79.0)
Islam	54 (13.5)	13 (6.5)	41 (20.5)
Christian	3 (0.7)	2 (1.0)	1 (0.5)
Education, n (%), *P*=.004			
High school or below	288 (71.8)	131 (65.2)	157 (78.5)
Bachelor or above	113 (28.2)	70 (34.8)	43 (21.5)
Food delivery company, n (%), *P*<.001			
A	188 (46.9)	102 (50.7)	86 (43.0)
B	148 (36.9)	38 (18.9)	110 (55.0)
C	60 (15.0)	56 (27.9)	4 (2.0)
D	5 (1.2)	5 (2.5)	0
Working duration as driver (months), *P*<.001			
Median (IQR)	12 (5-17)	8 (4-15)	12 (7-24)
Working hour per day, *P*=.004			
Median (IQR)	9 (8-10)	8 (7-10)	9 (8-12)
Number of daily food order deliveries, *P*<.001			
Median (IQR)	20 (15-25)	20 (20-30)	18 (15-20)
Daily income (THB[b,c]), *P*<.001			
Median (IQR)	600 (500-700), or US $15.90 ($13.25-$18.55)	600 (500-700), or US $15.90 ($13.25-$18.55)	500 (400-600), or US $13.25 ($10.60-$15.90)
Health insurance, n (%), *P*=.04			
Universal coverage	241 (60.1)	120 (59.7)	121 (60.5)
Civil servant	4 (1.0)	4 (2.0)	0
Social security	113 (28.2)	50 (24.9)	63 (31.5)
Self-pay	43 (10.7)	27 (13.4)	16 (8.0)
Pre-existing chronic illnesses, n (%), *P*=.66			
No	380 (94.8)	189 (94.0)	191 (95.5)
Yes	21 (5.2)	12 (6.0)	9 (4.5)
Cigarette smoker, n (%), *P*<.001			
Never smoke	245 (61.1)	124 (61.7)	121 (60.5)
Former smoker	59 (14.7)	42 (20.9)	17 (8.5)
Current smoker	97 (24.2)	35 (17.4)	62 (31.0)
Alcohol consumption, n (%), *P*=.99			
Never drink	272 (67.8)	136 (67.7)	136 (68.0)
Current drinker	129 (32.2)	65 (32.3)	64 (32.0)
COVID-19 HL[d] levels, n (%), *P*<.001			

Characteristics	Total (N=401)	Upper south (n=201)	Lower south (n=200)
Excellent	285 (71.1)	91 (45.3)	194 (97.0)
Moderate	98 (24.4)	92 (45.8)	6 (3.0)
Inadequate	18 (4.5)	18 (8.9)	0
COVID-19 preventive behavior levels, n (%), *P*<.001			
Excellent	355 (88.5)	165 (82.1)	190 (95.0)
Moderate	45 (11.2)	35 (17.4)	10 (5.0)
Inadequate	1 (0.3)	1 (0.5)	0

[a]*P* value: chi-square test for categorical outcomes and Mann-Whitney test for numerical outcomes.

[b]THB: Thai baht.

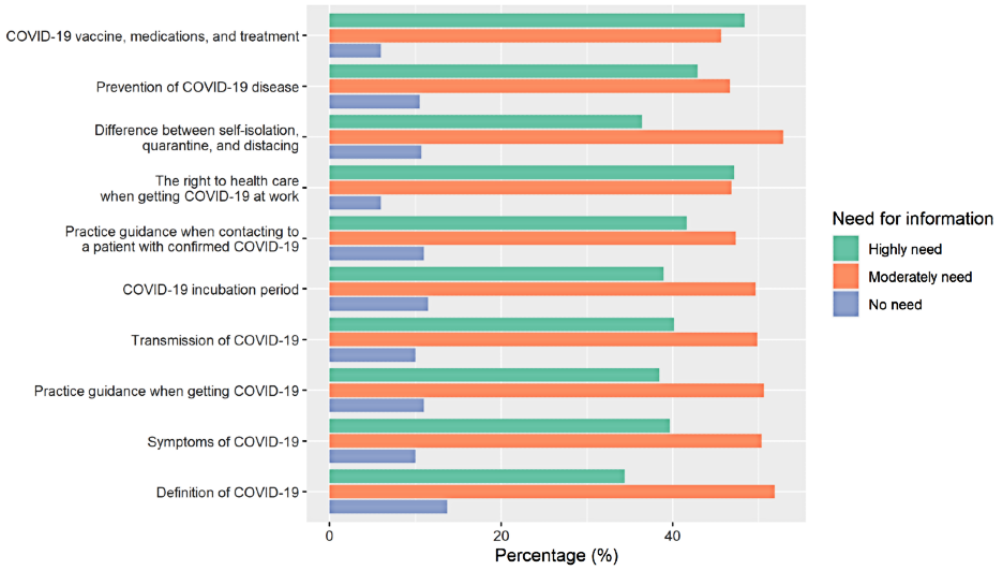
[c]An exchange rate of THB 1=US $0.026 has been applied.

[d]HL: health literacy.

Health Information Needs Regarding COVID-19

The participants' health information needs regarding COVID-19 are shown in Figure 1. The participants' most common health information needs regarding COVID-19 were related to vaccines, medications, and treatment (194/401, 48.4%), followed by the right to health care when getting COVID-19 at work (189/401, 47.1%). The least needed information was the definition of COVID-19 (55/401, 13.7%) and the COVID-19 incubation period (46/401, 11.5%).

Figure 1. Participants' COVID-19 information needs.



Health Literacy Toward COVID-19

Four main dimensions of HL regarding COVID-19 were identified: access to information, understanding, decision/judgment, and applying information, which included COVID-19 prevention behaviors. Regarding accessibility to health information related to COVID-19, the social media site Facebook was the most common source of information, followed by Line; 222 (55.4%) and 206 (51.4%) of 401 participants reported that they sometimes accessed information via websites and YouTube, respectively (Table 2), while 147 (36.7%) of the participants often accessed information from broadcasts, such as television. In addition, three-fourths of the participants revealed that they had heard of COVID-19 from relatives and colleagues. The information resources accessed by participants differed between the 2 regions of southern Thailand. The lower-south region had a significantly higher frequency of access to all resources, as presented in Table 2.

Of the 401 participants, 380 (94.8%) perceived that they had sufficient knowledge about COVID-19 (Table 3). A majority (393/401, 98%) obtained information and notifications regarding COVID-19 from their food delivery companies. Most participants (393/401, 98%) reported that they could easily access information about handwashing and mask wearing, although 326 (81.3%) participants were able to access that information less frequently than weekly. In addition, 380 (94.8%) participants usually verified the handwashing and mask-wearing information.

Table 2. Information resources regarding COVID-19 prevention.

Information resources and frequency of access	Total (N=401), n (%)	Upper south (n=201), n (%)	Lower south (n=200), n (%)
Website, *P*[a]<.001			
Frequent	45 (11.2)	36 (17.9)	9 (4.5)
Sometimes	222 (55.4)	97 (48.3)	125 (62.5)
Never	134 (33.4)	68 (33.8)	66 (33.0)
Line, *P*<.001			
Frequent	196 (48.9)	127 (63.2)	69 (34.5)
Sometimes	182 (45.4)	59 (29.3)	123 (61.5)
Never	23 (5.7)	15 (7.5)	8 (4.0)
Facebook, *P*=.01			
Frequent	299 (74.6)	140 (69.6)	159 (79.5)
Sometimes	86 (21.4)	48 (23.9)	38 (19.0)
Never	16 (4.0)	13 (6.5)	3 (1.5)
YouTube, *P*<.001			
Frequent	101 (25.2)	68 (33.8)	33 (16.5)
Sometimes	206 (51.4)	54 (26.9)	152 (76.0)
Never	94 (23.4)	79 (39.3)	15 (7.5)
Television, *P*<.001			
Frequent	147 (36.7)	17 (8.5)	130 (65.0)
Sometimes	120 (29.9)	59 (29.3)	61 (30.5)
Never	134 (33.4)	125 (62.2)	9 (4.5)
Relatives, *P*<.001			
Frequent	146 (36.4)	16 (8.0)	130 (65.0)
Sometimes	149 (37.2)	80 (39.8)	69 (34.5)
Never	106 (26.4)	105 (52.2)	1 (0.5)
Colleagues, *P*<.001			
Frequent	159 (39.7)	22 (10.9)	137 (68.5)
Sometimes	150 (37.4)	87 (43.3)	63 (31.5)
Never	92 (22.9)	92 (45.8)	0

[a]*P* value: chi-square test.

The understanding, judgment, and application of information regarding COVID-19 prevention were assessed using a set of questions. Participants' understanding of COVID-19 prevention is shown in Table 4. More than 90% (360/401) of the participants correctly answered 10 (83%) of the 12 questions. Interestingly, 157 (39.2%) drivers did not know that they should change their masks daily. Although the participants' ability to judge information and make decisions regarding COVID-19 prevention varied, most (362/401, 90.3%) participants reported that they were able to exchange information about COVID-19 prevention measures with health care providers and their colleagues, and 360 (89.8%) participants had decided to wear a mask and wash their hands every time they delivered food to a customer. In contrast, 132 (32.9%) drivers found it extremely difficult to question health care providers when they were confused about proper mask wearing and handwashing, as shown in Table 5.

Regarding the application of information, the majority (373/401, 93%) of the participants reported that they always keep a distance of at least 1 m from others when standing or sitting at a restaurant and when delivering food to customers. Almost all (348/401, 86.8%) participants monitored themselves and their families for COVID-19–related symptoms (Table 6). However, 69 (17.2%) participants did not use, or seldom used, a Food and Drug Administration (FDA)–certified mask. COVID-19 prevention behaviors are presented in Table 6. Most (395/401, 98.5%) participants reported that they always clean their hands using either regular soap and water or an alcohol-based hand rub before and after delivering food to consumers, and they wear a mask regularly while in public places, particularly when picking up and delivering food/beverage orders (390/401,

97.3%). Interestingly, 269 (67.1%) participants reported that instead of taking their masks off, they always pull them down under the chin to talk and to eat or drink.

Drivers' distributions according to the 3 dimensions (understanding, judgment, and application) are shown in Figure 2. Most participants had excellent HL levels in all 3 dimensions. Of the 401 participants, 345 (86%) had an excellent understanding of COVID-19, while 13 (3.2%) had an inadequate understanding. Excellent and moderate judgment accounted for 62.3% (250/401) and 28.9% (116/401) of the participants, respectively. Of the 401 participants, 309 (77.1%) were rated "excellent" at applying COVID-19 information to preventive behaviors, whereas 16 (4%) were rated "inadequate." A large proportion (285/401, 71.1%) of drivers had excellent HL, followed by moderate (98/401, 24.4%) and inadequate (18/401, 4.5%) HL. Most (355/401, 88.5%) of the participants had excellent COVID-19 prevention behavior. The difference in the level of each dimension of HL was statistically significant between participants in 2 different southern regions of Thailand. Participants in the lower south were more likely to have excellent HL and preventive behavior levels than those in the upper south (Table 1).

Table 3. Factors related accessibility to health information regarding COVID-19 prevention.

Variables	Total (N=401), n (%)	Upper south (n=201), n (%)	Lower south (n=200), n (%)
Perceived sufficient knowledge regarding COVID-19 prevention, *P*<.001			
Yes	380 (94.8)	181 (90.0)	199 (99.5)
No	21 (5.2)	20 (10.0)	1 (0.5)
Obtaining information regarding COVID-19 prevention from the company, *P*=.28			
Yes	393 (98.0)	195 (97.0)	198 (99.0)
No	8 (2.0)	6 (3.0)	2 (1.0)
Ability to access information regarding handwashing and mask wearing, *P*=.07			
Access	393 (98.0)	194 (96.5)	199 (99.5)
No access	8 (2.0)	7 (3.5)	1 (0.5)
Frequency of seeking information, *P*<.001			
Less than weekly	326 (82.1)	135 (68.2)	191 (96.0)
Weekly	71 (17.9)	63 (31.8)	8 (4.0)
Information verification, *P*=.13			
Yes	380 (95.7)	186 (93.9)	194 (97.5)
No	17 (4.3)	12 (6.1)	5 (2.5)

Table 4. Understanding of drivers (N=401) regarding COVID-19 prevention.

Variables	Yes, n (%)	No, n (%)
COVID-19 is an infectious disease caused by the SARS-CoV-2, which has been found since 2019.	380 (94.8)	21 (5.2)
The most common symptoms of COVID-19 include cough and tiredness.	385 (96.0)	16 (4.0)
Older adults, including those with diabetes/hypertension/heart diseases/chronic lung disease/cancers, are at the highest risk of COVID-19–related adverse outcomes and mortality.	382 (95.3)	19 (4.7)
COVID-19 patients with or without symptoms can pass on the virus.	384 (95.8)	17 (4.2)
If people feel sick or have trouble breathing, they should go to see a doctor immediately.	375 (93.5)	26 (6.5)
The spread of COVID-19 occurs via airborne particles and droplets when they exhale (eg, speaking, coughing, and sneezing).	382 (95.3)	19 (4.7)
COVID-19 infection is possible by touching contaminated surfaces (eg, doorknobs, handles, and tables).	382 (95.3)	19 (4.7)
People should be self-quarantined for 14 days after their last contact with infected patients and monitor themselves for fever, cough, etc.	390 (97.3)	11 (2.7)
Mask wearing can prevent the transmission of COVID-19.	361 (90.0)	40 (10.0)
Changing masks daily is effectively preventive behavior.	244 (60.8)	157 (39.2)
A cloth face mask should be washed at least once a day.	350 (87.3)	51 (12.7)
Handwashing following 7 steps and for at least 20 seconds is an effective process.	377 (94.0)	24 (6.0)

Table 5. Judgment of information toward COVID-19 prevention among drivers (N=401).

Variables	Extremely easy/slightly easy	Neutral	Extremely difficult/slightly difficult
Question health care providers when you are confused about proper mask wearing and handwashing.	243 (60.6)	26 (6.5)	132 (32.9)
Question about the effective ways of preventing novel coronavirus infection from health care providers and others.	251 (62.6)	22 (5.5)	128 (31.9)
Decide how to correctly wear a mask and wash hands when delivering food to consumer.	342 (85.3)	27 (6.7)	32 (8.0)
Decide to wear a mask and wash hands every time when delivering food to each consumer.	360 (89.8)	27 (6.7)	14 (3.5)
Know the advantages and disadvantages of each preventive measure when working as a food delivery driver.	331 (82.5)	45 (11.2)	25 (6.2)
Exchange information about COVID-19 prevention measures with health care providers, including colleagues at work.	362 (90.3)	28 (7.0)	11 (2.7)

Table 6. Application of information regarding COVID-19 prevention and preventive behaviors among drivers (N=401).

Variables	Often/always	Sometimes	Never/seldom
Application of information toward COVID-19 prevention			
Monitor yourself and your family for COVID-19–related symptoms (eg, fever, cough, trouble breathing) and suggest seeing a doctor immediately when getting these symptoms.	348 (86.8)	31 (7.7)	22 (5.5)
Consider using an appropriate mask by yourself when working as a food delivery driver.	341 (85.0)	31 (7.7)	29 (7.2)
Use a mask certified by the Food and Drug Administration (FDA).	292 (72.8)	40 (10.0)	69 (17.2)
Wash your hands following the 7 steps of effective handwashing.	345 (86.0)	34 (8.5)	22 (5.5)
Keep a distance of at least 1 m from others when standing or sitting at a restaurant and delivering food to consumers.	373 (93.0)	19 (4.7)	9 (2.3)
Preventive behaviors toward COVID-19			
Replace a mask with a new one once it is wet or soiled from saliva or mucus.	384 (95.8)	10 (2.5)	7 (1.7)
Pull the mask down under the chin to talk to and to eat or drink instead of taking the mask off.	269 (67.1)	35 (8.7)	97 (24.2)
Wash your hands before touching your face, eyes, nose, or mouth.	349 (87.0)	42 (10.5)	10 (2.5)
Clean your hands immediately with soap and water or an alcohol-based hand rub after touching doorknobs, handrails, light switches, and much more.	379 (94.5)	19 (4.7)	3 (0.8)
Wear a mask regularly when in public places (eg, market, restaurant), particularly when picking up and delivering food/beverage orders.	390 (97.3)	9 (2.2)	2 (0.5)
Clean your hands regularly with soap and water or an alcohol-based hand rub before and after delivering food to consumers.	395 (98.5)	4 (1.0)	2 (0.5)

The Pearson moment correlation coefficient (r) was used to examine the relationship between participants' age, work duration (months), number of working hours per day, understanding, judgment, application, and behavior scores. Figure 3 shows weak-to-strong correlation coefficients for the predictor variables associated with the HL of participants. The relationships between judgment and HL and between application and HL were positively correlated and statistically significant (r=0.83 and 0.77, respectively); a statistically significant moderate positive correlation was found between understanding and HL (r=0.45). These results suggest that higher scores on participants' judgment and application of information are associated with higher HL scores. Furthermore, a significant moderate positive correlation was found between participants' application of information and behavior toward COVID-19 prevention (r=0.38), whereas a weak positive correlation was found between HL and behavior (r=0.22). Meanwhile, age was significantly weakly positively correlated with work duration (r=0.29), understanding (r=0.25), and HL (r=0.26).

Associations between factors, HL, and behavior toward COVID-19 prevention were determined using univariate and multivariate analyses and are presented in Table 7. HL and behavior scores were treated as continuous variables; subsequently, they were analyzed using 2 models, HL and behavior models. All significant factors from the univariate analysis were included in the multivariate analysis in the HL model. Results from the multivariate analysis showed factors significantly associated with HL regarding COVID-19, which were the region participants worked in, health insurance status (social security scheme), information resource access (YouTube, television, and relatives), knowledge of COVID-19, and

frequency of seeking information. Participants in the lower-south region of Thailand had higher HL than those in the upper-south region (P<.001). Participants who frequently accessed information via television, and had knowledge of COVID-19, had significantly higher HL than the others. In contrast, participants who belonged to self-payment health care, those who accessed information through YouTube and relatives, and those who frequently accessed information less than weekly had significantly lower HL than the others.

For a multivariable model of COVID-19 prevention behavior, the region participants worked in, their sex, number of working hours per day, information resources (Facebook), information verification, and HL score were significantly associated with their COVID-19 prevention behavior. The COVID-19 prevention behavior of men was statistically significantly lower than that of women. Participants who accessed information through Facebook had higher COVID-19 prevention behavior than those who never accessed it; those who verified information before its application also had higher behavior than those who did not verify information. Additionally, higher HL was significantly associated with higher COVID-19 prevention behaviors.

Figure 2. Participants' health literacy levels on each dimension.

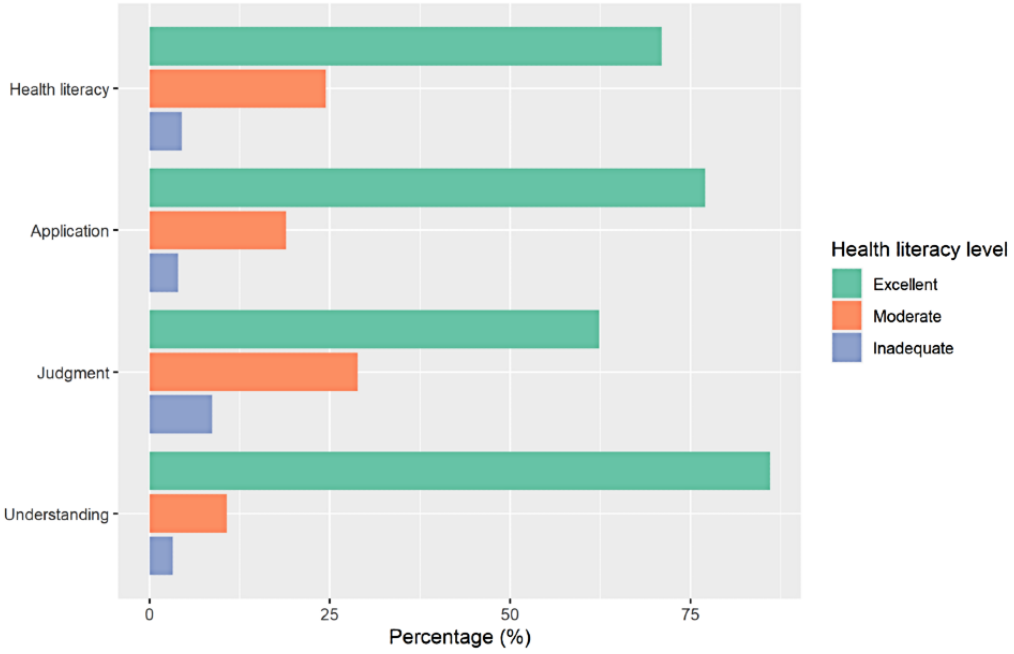


Figure 3. Correlation matrix of research variables.

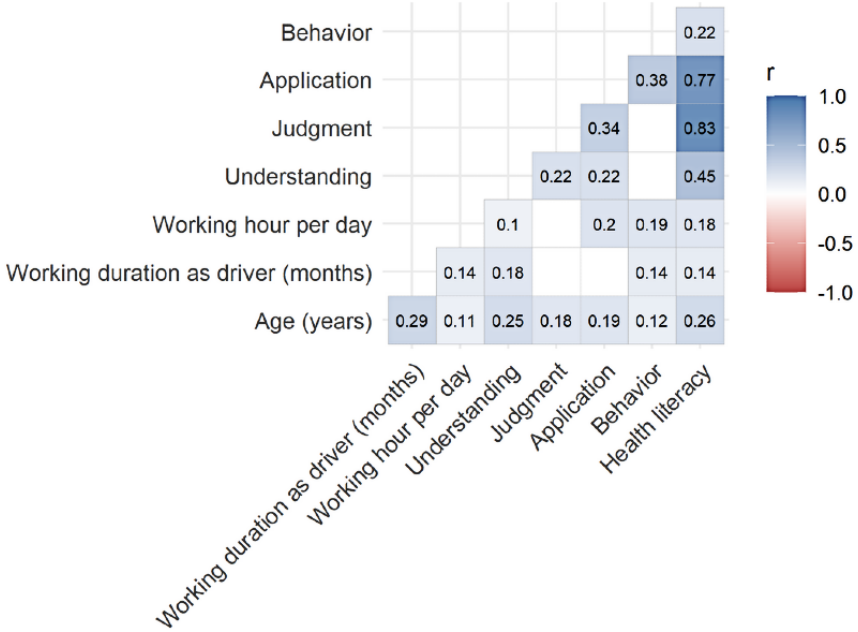


Table 7. Univariate and multivariate regression analyses demonstrating associations between independent variables and HL score and behavior score regarding COVID-19 prevention.

Characteristic	HL[a] score univariate analysis		HL score multivariate analysis		Behavior score univariate analysis		Behavior score multivariate analysis	
	Crude coefficient (95% CI)	*P* value	Adjusted coefficient[b] (95% CI)	*P* value	Crude coefficient (95% CI)	*P* value	Adjusted coefficient[b] (95% CI)	*P* value
Region (lower south vs upper south)	6.44 (5.26-7.62)	<.001	7.03 (5.36-8.7)	<.001	1.36 (0.79-1.92)	<.001	0.81 (0.17-1.45)	.01
Sex (female vs male)	0.99 (–0.51 to 2.49)	.20	N/A[c]	N/A	–0.42 (–1.07 to 0.23)	.20	–0.76 (–1.38 to –0.13)	.02
Age (years)	0.18 (0.11-0.24)	<.001	N/A	N/A	0.03 (0.01-0.06)	.01	N/A	N/A
Religion (reference=Buddhism)								
Islam	0.95 (–1.02 to 2.91)	.34	N/A	N/A	0.35 (–0.5 to 1.21)	.40	N/A	N/A
Christian	5.75 (–1.98 to 13.47)	.14	N/A	N/A	–0.31 (–3.67 to 3.05)	.86	N/A	N/A
Education (bachelor or above vs high school or below)	–2.00 (–3.48 to –0.53)	.01	N/A	N/A	0.15 (–0.5 to 0.79)	.66	N/A	N/A
Food delivery company (reference=company A)								
B	1.64 (0.18-3.1)	.03	N/A	N/A	1.03 (0.4-1.66)	.001	N/A	N/A
C	–0.87 (–2.86 to 1.13)	.39	N/A	N/A	–0.01 (–0.87 to 0.85)	.99	N/A	N/A
D	–1.42 (–7.43 to 4.58)	.64	N/A	N/A	–0.92 (–3.51 to 1.67)	.49	N/A	N/A
Working duration as driver (months)	0.11 (0.03-0.19)	.01	N/A	N/A	0.05 (0.01-0.08)	.01	N/A	N/A
Working hour per day	0.46 (0.2-0.71)	<.001	N/A	N/A	0.22 (0.11-0.33)	<.001	0.17 (0.06-0.28)	.002
Number of daily food order deliveries	–0.11 (–0.21 to –0.02)	.02	N/A	N/A	–0.02 (–0.07 to 0.02)	.24	N/A	N/A
Daily income (THB[d])	0 (–0.01 to 0)	.16	N/A	N/A	0 (0-0)	.99	N/A	N/A
Pre-existing chronic illnesses (yes vs no)	–0.35 (–3.34 to 2.65)	.82	N/A	N/A	–0.78 (–2.08 to 0.52)	.24	N/A	N/A
Cigarette smoker (reference=never smoke)								
Former smoker	–1.33 (–3.26 to 0.61)	.18	N/A	N/A	–0.23 (–1.08 to 0.61)	.59	N/A	N/A
Current smoker	1.77 (0.17-3.37)	.03	N/A	N/A	0.40 (–0.29 to 1.1)	.26	N/A	N/A
Alcohol consumption (current drinker vs never drink)	0.39 (–1.05 to 1.82)	.60	N/A	N/A	0.15 (–0.47 to 0.78)	.63	N/A	N/A
Health insurance status (reference=universal coverage)								
Civil servant	–6.2 (–12.75 to 0.35)	.06	–1.45 (–6.63 to 3.73)	.58	–0.54 (–3.47 to 2.38)	.71	N/A	N/A
Social security	–0.04 (–1.53 to 1.45)	.96	0.49 (–0.69 to 1.67)	.41	–0.12 (–0.79 to 0.54)	.71	N/A	N/A
Self-pay	–4.96 (–7.11 to –2.8)	<.001	–2.25 (–3.98 to –0.53)	.01	–0.58 (–1.54 to 0.38)	.23	N/A	N/A
Website (reference=never)								
Frequent	0.03 (–2.3 to 2.36)	.98	N/A	N/A	–0.46 (–1.47 to 0.54)	.37	N/A	N/A
Sometimes	0.05 (–1.42 to 1.52)	.95	N/A	N/A	0.46 (–0.17 to 1.09)	.15	N/A	N/A
Line (reference=never)								

Characteristic	HL[a] score univariate analysis		HL score multivariate analysis		Behavior score univariate analysis		Behavior score multivariate analysis	
	Crude coefficient (95% CI)	*P* value	Adjusted coefficient[b] (95% CI)	*P* value	Crude coefficient (95% CI)	*P* value	Adjusted coefficient[b] (95% CI)	*P* value
Frequent	1.31 (–1.68 to 4.29)	.39	N/A	N/A	–0.01 (–1.31 to 1.29)	.99	N/A	N/A
Sometimes	2.69 (–0.3 to 5.69)	.08	N/A	N/A	0.43 (–0.88 to 1.73)	.52	N/A	N/A
Facebook (reference=never)								
Frequent	0.17 (–3.25 to 3.58)	.92	N/A	N/A	1.92 (0.44-3.39)	.01	1.66 (0.24-3.08)	.02
Sometimes	–1.35 (–4.98 to 2.29)	.47	N/A	N/A	1.50 (–0.06 to 3.07)	.06	1.35 (–0.15 to 2.86)	.08
YouTube (reference=never)								
Frequent	–3.09 (–4.99 to –1.2)	.001	–2.17 (–3.88 to –0.47)	.01	0.38 (–0.45 to 1.21)	.37	N/A	N/A
Sometimes	–0.53 (–2.19 to 1.12)	.53	–4.08 (–5.72 to –2.45)	<.001	0.63 (–0.1 to 1.35)	.09	N/A	N/A
Television (reference=never)								
Frequent	5.58 (4.08-7.08)	<.001	2.81 (0.95-4.66)	.003	1.29 (0.61-1.97)	<.001	N/A	N/A
Sometimes	1.18 (–0.4 to 2.75)	.14	1.54 (–0.11 to 3.19)	.07	0.40 (–0.32 to 1.13)	.27	N/A	N/A
Relatives (reference=never)								
Frequent	3.03 (1.42 to 4.64)	<.001	–4.19 (–6.25 to –2.13)	<.001	0.86 (0.12-1.6)	.02	N/A	N/A
Sometimes	–2.54 (–4.14 to –0.93)	.002	–5.36 (–7.1 to –3.63)	<.001	–0.07 (–0.8 to 0.67)	.86	N/A	N/A
Colleagues (reference=never)								
Frequent	2.72 (1.02-4.42)	.002	N/A	N/A	1.22 (0.47-1.97)	.002	N/A	N/A
Sometimes	–1.43 (–3.14 to 0.29)	.10	N/A	N/A	0.15 (–0.61 to 0.91)	.69	N/A	N/A
Perceived sufficient knowledge of COVID-19 (yes vs no)	5.35 (2.25-8.44)	<.001	4.61 (2.1-7.12)	<.001	0.72 (–0.63 to 2.08)	.30	N/A	N/A
Obtaining information from the company (yes vs no)	3.33 (–2.67 to 9.33)	.28	N/A	N/A	0.90 (–1.7 to 3.51)	.50	N/A	N/A
Accessibility to information (access vs no access)	5.83 (1.1-10.57)	.02	2.89 (–0.81 to 6.59)	.13	1.21 (–0.85 to 3.28)	.25	N/A	N/A
Frequency of seeking information (weekly vs less than weekly)	–6.59 (–8.22 to –4.97)	<.001	–3.04 (–4.56 to –1.52)	<.001	–1.13 (–1.88 to –0.38)	.003	N/A	N/A
Information verification (yes vs no)	1.82 (–1.49 to 5.12)	.28	N/A	N/A	2.01 (0.59-3.43)	.01	1.42 (0.05-2.79)	.04
HL score	N/A	N/A	N/A	N/A	0.10 (0.05-0.14)	<.001	0.06 (0.01-0.1)	.02

[a]HL: health literacy.

[b]Adjusted by the backward stepwise method.

[c]N/A: not applicable.

[d]THB: Thai baht.

Discussion

Principal Findings

Food delivery drivers are at high risk of exposure to SARS-CoV-2 infection at work, and those who are suspected of contracting the infection might play a role in actively transmitting the infection to customers [12,15]. This study is the first cross-sectional survey on comprehensive HL related to COVID-19 prevention among food delivery drivers in southern Thailand. We aimed to explore drivers' information needs and resources and assess their HL levels regarding COVID-19 prevention.

The findings of this study revealed that almost half of food delivery drivers need to know about vaccines, medications, treatment for COVID-19, as well as the right to health care when getting SARS-CoV-2 infection. Participants were less concerned about the natural history of COVID-19 and its definition. Various studies conducted among medical learners and undergraduate and graduate students in Jordan [26,27], senior pharmacy students at the British University in Egypt [28], and internet users and residents in China [29,30] have revealed that participants use social media platforms as their primary source of information about COVID-19. These results were consistent with our findings that social media, such as Facebook and Line, was the most common source of information regarding COVID-19 prevention used by drivers. These sources of information tend to be common for every aspect of daily life during working hours. However, health information from social media platforms may only provide broad and nonspecific information about COVID-19 prevention. Drivers may need specific and customized information related to their lifestyle, especially information regarding the right to health care when getting a SARS-CoV-2 infection, which they can get from local health care workers and their companies through social media platforms with user-friendly content, for instance, eye-catching images that separate text. More importantly, Thai government authorities and food delivery companies should work more on clarifying the rights and responsibilities of drivers when they are at high risk of COVID-19 infection during working hours by improving relevant policies and regulations. Building effective daily routine online communication for drivers is key to successfully increasing their HL levels and preventive behaviors. A closed-class Facebook group can be used as an alternative learning support application regarding COVID-19 prevention because it easily retrieves information sources and shares them with colleagues for intellectual discussion [31].

Our study also found that relatives and colleagues played a role in disseminating information to the drivers. However, health workers have been reported to be a source that can provide accurate information in an institutional-based cross-sectional study among university students in Colombia [32]. The majority of drivers in our study showed that they usually obtained and verified information through any resource, particularly social media. A previous review by González-Padilla and Tortolero-Blanco [33] presented the advantages and disadvantages associated with the use of social media platforms during the pandemic. Important disadvantages are that invalid or outdated information is common, and information on social media platforms is often not fact-checked. People should be aware of these disadvantages when seeking information via social media. An important advantage is the rapid dissemination of educational content. Thus, effective communication via accessible information resources, such as social media platforms, including the frequency of health information delivery, may help improve drivers' understanding of COVID-19 prevention.

The 4 dimensions of HL are competencies in health information processing, including accessing, understanding, decision-making/judging, and applying information. These overall competencies subsequently influence individual health behaviors [19]. HL is an important factor in improving the understanding, risk awareness, and decision-making regarding preventive behaviors and lifestyles during a pandemic [34,35]. We did not find any studies that have reported the status of HL regarding COVID-19 prevention among food delivery drivers. Most drivers in this study had excellent HL regarding COVID-19 prevention, with only a few having inadequate HL. Various researchers have indicated that better health outcomes are associated with higher HL, which results from the acquisition of new knowledge; higher HL is also associated with more positive attitudes, greater self-efficacy, and positive health behaviors [18]. We found that it was possible to develop the HL of drivers from the beginning, and the drivers accessed information regarding COVID-19 prevention often, mostly via social media. The participants had excellent levels of understanding, judgment, and application of information regarding COVID-19 prevention, which translates into excellent HL. In the multivariate analysis, drivers in the lower south of Thailand were more likely to have excellent HL levels than those in the upper south. Food delivery services were established in the lower-south region of Thailand in 2019 and started in the upper-south region a year later. During the study period, the Thai government declared a national state of emergency in an intensified attempt to stun the spread of the coronavirus. Thailand's Center for COVID-19 Situation Administration (CCSA) categorizes provinces into dark-red, red, and orange zones in decreasing order of the strictness of social and public health measures. The lower-south region of Thailand was categorized as dark red because of the rapid transmission of COVID-19 and cluster infections. This region exercised social distancing, lockdowns, and a nighttime curfew, as prescribed through concrete government interventions. Food delivery drivers needed to work according to strict regulations during the pandemic to mitigate their risk of SARS-CoV-2 infection [36]. Public health promotion campaigns have also been established in the lower south of Thailand, including a COVID-19 vaccination campaign. The regulations and campaigns in the lower-south region may have led to a higher HL level among drivers in the lower south than among those who work in the upper south of Thailand.

We found that the frequency of seeking information and information resources, such as YouTube and relatives, is negatively associated with HL among drivers. These findings were in contrast to those from a survey of Japanese people that showed that higher literacy is positively associated with health information access and obtaining sufficient information from

multiple sources [37]. These findings were also in contrast to a cross-sectional study among Thai older adults that showed that the more access older adults have to health information, the higher is their HL [38]. However, not only informational accessibility skills but also the ability to obtain credible health information determines individuals' HL [19,37,39]. The first non-German replication of the Ebbinghaus forgetting experiment revealed individual differences in the lengths of memory retention intervals [40]. The Ebbinghaus forgetting curve describes forgetting over intervals ranging from 20 minutes to 31 days after information is accessed. It is a key psychology study that has affected our understanding of information literacy. According to the HL framework, some demographic factors, such as age, moderate the development of HL; however, increasing age alone could not possibly contribute to the higher HL in our study, and this result might be confounded by other factors in the multivariate analysis. In addition, participants with perceived sufficient knowledge regarding COVID-19 in this study may have higher HL, as there are reports indicating that knowledge mediates the effects of HL [41].

Regarding behaviors about COVID-19 prevention, this study found that overall HL has a weak positive correlation with drivers' COVID-19 prevention behavior. No evidence of a correlation between judgment and application of information and COVID-19 prevention behavior was found. The finding on the positive correlation between HL and COVID-19 prevention behaviors was similar to an online survey of Vietnamese health care workers by Do et al [42], who reported a significant positive association between HL and health care workers' self-reported adherence to occupational infection prevention and control measures regarding COVID-19. In a national web-based cross-sectional survey of Chinese internet users, Li and Liu [29] also revealed that HL is significantly associated with the self-reported practice of protective behaviors against COVID-19 during the pandemic.

In a multivariate analysis to predict COVID-19 prevention behavior, we found that using Facebook to access information is a factor associated with participants' preventive behavior level. Drivers who frequently accessed information via Facebook were more likely to exhibit excellent behavior than those who used Facebook only occasionally. Our finding was consistent with a national web-based cross-sectional survey of Chinese internet users authored by Li and Liu [29], who reported that social media use frequency significantly predicts COVID-19 prevention behaviors. It has also been reported that social media use plays a positive role in individual health behaviors [43].

Health information shared on social media has the potential to improve people's preventive behavior toward COVID-19. However, information access skills and the ability to obtain credible health information determine individuals' HL [19,37,39]. Library skills and internet skills, including information evaluation skills, were found to be the most essential skills related to health behaviors, and information verification is also an important skill to confirm the credibility of information before applying it in practice [44,45]. This is consistent with the finding of this study that drivers who verified information were more likely to have good COVID-19 health behaviors. It is crucial to ensure that individuals have the necessary skills to obtain and judge health information before they apply it in practice. In addition, the HL score was a factor associated with preventive behavior, and drivers with higher HL scores were more likely to have good COVID-19 prevention behaviors.

Strengths and Limitations

The major strength of this study is that it was the first study conducted among Thai food delivery drivers in southern Thailand during the COVID-19 pandemic, exploring the drivers' HL status and its associated factors.

This study also had some limitations. First, we used convenience sampling to recruit eligible drivers to join the study. That drivers who were interested in this study might share similar traits (eg, being in a younger population group or working in the same food delivery company) that could result in bias cannot be ruled out. Second, the results from this study may not be applicable to drivers in other regions of Thailand because of the different provincial zones categorized by the CCSA with their own specific measures to stun the spread of coronavirus. Future research conducted with a bigger sample size would likely improve the generalizability of the results. Finally, our data collection process overlapped with the COVID-19 pandemic and was performed in the late afternoon until the evening. Thus, only drivers who were available at the time participated in the study. This might have affected the external validity of this study as drivers who did not participate in the study might have had a different socioeconomic status or different HL and preventive behaviors. Future studies could use different means to distribute the survey to a more diverse audience.

Conclusion

Most food delivery drivers reported sufficient knowledge about COVID-19. Most drivers can easily access information about handwashing and mask wearing. They were able to frequently access information and usually verified information to confirm its credibility. They had excellent HL and COVID-19 prevention behaviors. Most had good levels of understanding, judgment, and application of information regarding COVID-19. As good understanding and decision-making/judgment of information may not reflect good preventive behaviors, conventional health education alone may not be effective for COVID-19 prevention among drivers. Thus, an effective technique to apply the information to practice would enhance drivers' HL, including COVID-19 prevention behavior. Many risk factors were observed. Information access skills alone may not determine individuals' HL, but the ability to obtain credible health information, effective interactive communication, and knowledge could determine HL. To ensure the long-term prevention of COVID-19 in the food delivery sector, action must be taken now to eliminate the hidden barriers to health information communication with the drivers in the pandemic context. For practitioners as drivers, it is important to have sufficient support from restaurant staff, consumers, food delivery companies, and government authorities to ensure contact-free delivery and strict use of new face masks, gloves, and hand sanitizers for COVID-19 risk mitigation. Policymakers should raise awareness of safety practices for drivers through effective techniques to ensure the successful application of health

information to practice, for instance, hands-on education and coaching with online learning regarding COVID-19 prevention that fits drivers' work baseline literacy. Importantly, building trust is key for successful collaboration between drivers and government authorities to enhance the drivers' HL and preventive behaviors regarding COVID-19. A better understanding of HL among food delivery drivers would be useful for planning effective interventions for this population, especially for Thai food delivery services.

Acknowledgments

This study was funded by the Individual Research Grant from Walailak University, Thailand (WU-IRG-64-049). Our appreciation is extended to the university for funding and supporting us throughout the duration of the project. We would like to thank Editage for English language editing. We would also like to express our appreciation for the hard work of the site research assistants in data collection. The authors also acknowledge all the study participants who took part in the study.

Authors' Contributions

KJ contributed to the study design and planning, data collection, statistical analysis, data interpretation, and writing of the manuscript. CT contributed to the study design and planning and revision of the manuscript. All the authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. Novel Coronavirus – China. 2020. URL: https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ [accessed 2021-03-02]
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
3. Milne GJ, Xie S, Poklepovich D, O'Halloran D, Yap M, Whyatt D. A modelling analysis of the effectiveness of second wave COVID-19 response strategies in Australia. Sci Rep 2021 Jun 07;11(1):11958 [FREE Full text] [doi: 10.1038/s41598-021-91418-6] [Medline: 34099788]
4. Dsouza D, Sharma D. Online food delivery portals during COVID-19 times: an analysis of changing consumer behavior and expectations. IJIS 2020 Dec 30;13(2):218-232. [doi: 10.1108/ijis-10-2020-0184]
5. Li C, Mirosa M, Bremer P. Review of online food delivery platforms and their impacts on sustainability. Sustainability 2020 Jul 08;12(14):5528. [doi: 10.3390/su12145528]
6. Zanetta LD, Hakim MP, Gastaldi GB, Seabra LMJ, Rolim PM, Nascimento LGP, et al. The use of food delivery apps during the COVID-19 pandemic in Brazil: the role of solidarity, perceived risk, and regional aspects. Food Res Int 2021 Nov;149:110671 [FREE Full text] [doi: 10.1016/j.foodres.2021.110671] [Medline: 34600673]
7. Chang H, Meyerhoefer C. COVID-19 and the demand for online food shopping services: empirical evidence from Taiwan. Am J Agric Econ 2020 Nov 23;103(2):448-465 [FREE Full text] [doi: 10.1111/ajae.12170]
8. Byrd K, Her E, Fan A, Almanza B, Liu Y, Leitch S. Restaurants and COVID-19: what are consumers' risk perceptions about restaurant food and its packaging during the pandemic? Int J Hosp Manag 2021 Apr;94:102821 [FREE Full text] [doi: 10.1016/j.ijhm.2020.102821] [Medline: 34866742]
9. Ortiz-Prado E, Henriquez-Trujillo AR, Rivera-Olivero IA, Lozada T, Garcia-Bereguiain MA, UDLA-COVID-19 team. High prevalence of SARS-CoV-2 infection among food delivery riders. A case study from Quito, Ecuador. Sci Total Environ 2021 May 20;770:145225 [FREE Full text] [doi: 10.1016/j.scitotenv.2021.145225] [Medline: 33513511]
10. Felicia. Thailand's Best Food Delivery Service. 2021. URL: https://thethaiger.com/guides/best-of/thailands-top-food-delivery-service [accessed 2021-07-19]
11. Sirikeratikul S. Thailand Online Food Delivery Market. 2020. URL: https://tinyurl.com/mwupb3dh [accessed 2022-10-10]
12. Burdorf A, Porru F, Rugulies R. The COVID-19 (coronavirus) pandemic: consequences for occupational health. Scand J Work Environ Health 2020 Mar 31;46(3):229-230. [doi: 10.5271/sjweh.3893]
13. Nguyen THD, Vu DC. Food delivery service during social distancing: proactively preventing or potentially spreading coronavirus disease-2019? Disaster Med Public Health Prep 2020 Jun;14(3):e9-e10 [FREE Full text] [doi: 10.1017/dmp.2020.135] [Medline: 32367789]
14. Liu C, Cao S. Beijing Deliveryman Who Sends 50 Orders per Day Confirmed with COVID-19. URL: https://www.globaltimes.cn/content/1192565.shtml [accessed 2021-03-02]
15. Nguyen THD, Vu DC. The largest epicenter of the coronavirus outbreak in Vietnam. Infect Control Hosp Epidemiol 2020 Aug;41(8):984-985 [FREE Full text] [doi: 10.1017/ice.2020.128] [Medline: 32279697]

16. Hindustan Times. Food Delivery Agent Tests Covid-19 +ve in Delhi, 72 Families Quarantined. 2020. URL: https://tinyurl.com/2dbe62wz [accessed 2022-10-10]
17. Centers for Disease Control and Prevention. What Food and Grocery Pick-Up and Delivery Drivers Need to Know about COVID-19. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/organizations/food-grocery-drivers.html [accessed 2021-03-02]
18. Baker DW. The meaning and the measure of health literacy. J Gen Intern Med 2006 Aug;21(8):878-883 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00540.x] [Medline: 16881951]
19. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
20. Liu H, Zeng H, Shen Y, Zhang F, Sharma M, Lai W, et al. Assessment tools for health literacy among the general population: a systematic review. Int J Environ Res Public Health 2018 Aug 10;15(8):1711 [FREE Full text] [doi: 10.3390/ijerph15081711] [Medline: 30103386]
21. Klamklay J, Sungkhapong A, Yodpijit NE, E. Patterson P. Anthropometry of the southern Thai population. Int J Ind Ergon 2008 Jan;38(1):111-118 [FREE Full text] [doi: 10.1016/j.ergon.2007.09.001]
22. Situation Awareness Team (SAT), Department of Disease Control. COVID-19 Situation in Southern Thailand. 2022. URL: https://ddc.moph.go.th/covid19-dashboard/?dashboard=province [accessed 2022-10-10]
23. Harvard Humanitarian Initiative. KoBo Toolbox. 2021. URL: http://www.kobotoolbox.org [accessed 2021-06-20]
24. Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. Gastroenterol Hepatol Bed Bench 2013;6(1):14-17 [FREE Full text] [Medline: 24834239]
25. Marques S, Lemos S. Health literacy assessment instruments: literature review. Audiol Commun Res 2017;22:e1757. [doi: 10.1590/2317-6431-2016-1757]
26. Khasawneh AI, Humeidan AA, Alsulaiman JW, Bloukh S, Ramadan M, Al-Shatanawi TN, et al. Medical students and covid-19: knowledge, attitudes, and precautionary measures. a descriptive study from Jordan. Front Public Health 2020;8:253 [FREE Full text] [doi: 10.3389/fpubh.2020.00253] [Medline: 32574313]
27. Olaimat AN, Aolymat I, Shahbaz HM, Holley RA. Knowledge and information sources about covid-19 among university students in Jordan: a cross-sectional study. Front Public Health 2020 May 29;8:254 [FREE Full text] [doi: 10.3389/fpubh.2020.00254] [Medline: 32574314]
28. Hamza MS, Badary OA, Elmazar MM. Cross-sectional study on awareness and knowledge of covid-19 among senior pharmacy students. J Community Health 2021 Feb;46(1):139-146 [FREE Full text] [doi: 10.1007/s10900-020-00859-z] [Medline: 32542552]
29. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the covid-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
30. Yang K, Liu H, Ma L, Wang S, Tian Y, Zhang F, et al. Knowledge, attitude and practice of residents in the prevention and control of COVID-19: an online questionnaire survey. J Adv Nurs 2021 Apr;77(4):1839-1855 [FREE Full text] [doi: 10.1111/jan.14718] [Medline: 33259651]
31. Ulla M, Perales W. Facebook as an integrated online learning support application during the COVID19 pandemic: Thai university students' experiences and perspectives. Heliyon 2021 Nov;7(11):e08317 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e08317] [Medline: 34746477]
32. Archila PA, Danies G, Molina J, Truscott de Mejía AM, Restrepo S. Towards covid-19 literacy: investigating the literacy levels of university students in Colombia. Sci Educ (Dordr) 2021;30(4):785-808 [FREE Full text] [doi: 10.1007/s11191-021-00222-1] [Medline: 33897108]
33. González-Padilla DA, Tortolero-Blanco L. Social media influence in the covid-19 pandemic. Int Braz J Urol 2020 Jul;46(suppl.1):120-124 [FREE Full text] [doi: 10.1590/S1677-5538.IBJU.2020.S121] [Medline: 32550706]
34. Mackenbach JP. The persistence of health inequalities in modern welfare states: the explanation of a paradox. Soc Sci Med 2012 Aug;75(4):761-769. [doi: 10.1016/j.socscimed.2012.02.031] [Medline: 22475407]
35. Okan O, Bollweg TM, Berens E, Hurrelmann K, Bauer U, Schaeffer D. Coronavirus-related health literacy: a cross-sectional study in adults during the COVID-19 infodemic in Germany. Int J Environ Res Public Health 2020 Jul 30;17(15):5503 [FREE Full text] [doi: 10.3390/ijerph17155503] [Medline: 32751484]
36. Ministry of Foreign Affairs, Kingdom of Thailand. Regulation Issued under Section 9 of the Emergency Decree on Public Administration in Emergency Situations B.E. 2548 (2005) (No. 22). 2021. URL: https://www.mfa.go.th/en/content/regulation-issued-under-section-9-emergency-2?page=5f22514b78568958aa0d5b85&menu=5d5bd3cb15e39c306002a9b9 [accessed 2022-10-10]
37. Suka M, Odajima T, Okamoto M, Sumitani M, Igarashi A, Ishikawa H, et al. Relationship between health literacy, health information access, health behavior, and health status in Japanese people. Patient Educ Couns 2015 May;98(5):660-668 [FREE Full text] [doi: 10.1016/j.pec.2015.02.013] [Medline: 25739344]

38. Pechrapa K, Yodmai K, Kittipichai W, Charupoonpol P, Suksatan W. Health literacy among older adults during covid-19 pandemic: a cross-sectional study in an urban community in Thailand. Ann Geriatr Med Res 2021 Dec;25(4):309-317 [FREE Full text] [doi: 10.4235/agmr.21.0090] [Medline: 34735760]
39. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
40. Murre JMJ, Dros J. Replication and analysis of Ebbinghaus' forgetting curve. PLoS One 2015;10(7):e0120644 [FREE Full text] [doi: 10.1371/journal.pone.0120644] [Medline: 26148023]
41. Lee S, Arozullah A, Cho Y. Health literacy, social support, and health: a research agenda. Soc Sci Med 2004 Apr;58(7):1309-1321 [FREE Full text] [doi: 10.1016/s0277-9536(03)00329-0]
42. Do BN, Tran TV, Phan DT, Nguyen HC, Nguyen TTP, Nguyen HC, et al. Health literacy, eHealth literacy, adherence to infection prevention and control procedures, lifestyle changes, and suspected covid-19 symptoms among health care workers during lockdown: online survey. J Med Internet Res 2020 Nov 12;22(11):e22894 [FREE Full text] [doi: 10.2196/22894] [Medline: 33122164]
43. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
44. Ivanitskaya LV, Hanisko KA, Garrison JA, Janson SJ, Vibbert D. Developing health information literacy: a needs analysis from the perspective of preprofessional health students. J Med Libr Assoc 2012 Oct;100(4):277-283 [FREE Full text] [doi: 10.3163/1536-5050.100.4.009] [Medline: 23133327]
45. Shehata A. Health information behaviour during COVID-19 outbreak among Egyptian library and information science undergraduate students. Inf Dev 2020 Dec 07;37(3):417-430. [doi: 10.1177/0266666920976181]

Abbreviations

CCSA: Center for COVID-19 Situation Administration
HL: health literacy
THB: Thai baht

Edited by A Kushniruk; submitted 03.03.22; peer-reviewed by W Suksatan, WC Poon; comments to author 02.08.22; revised version received 02.09.22; accepted 06.10.22; published 13.10.22.

Please cite as:
Jandee K, Thanapop C
Food Delivery Drivers' Health Literacy Regarding COVID-19 Prevention and Protective Behaviors During the COVID-19 Pandemic: Cross-sectional Survey in Southern Thailand
JMIR Hum Factors 2022;9(4):e37693
URL: https://humanfactors.jmir.org/2022/4/e37693
doi:10.2196/37693
PMID:36227652

Original Paper

Understanding Human Factors Challenges on the Front Lines of Mass COVID-19 Vaccination Clinics: Human Systems Modeling Study

Ryan Tennant[1*], BASc, MASc; Moses Tetui[2,3*], PhD; Kelly Grindrod[2*], BScPharm, MSc, PharmD; Catherine M Burns[1*], PhD, PEng

[1]Department of Systems Design Engineering, Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada

[2]School of Pharmacy, University of Waterloo, Kitchener, ON, Canada

[3]Department of Epidemiology and Global Health, Umea University, Umea, Sweden

[*]all authors contributed equally

Corresponding Author:
Ryan Tennant, BASc, MASc
Department of Systems Design Engineering
Faculty of Engineering
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567
Email: drtennan@uwaterloo.ca

Abstract

Background: Implementing mass vaccination clinics for COVID-19 immunization has been a successful public health activity worldwide. However, this tightly coupled system has many logistical challenges, leading to increased workplace stress, as evidenced throughout the pandemic. The complexities of mass vaccination clinics that combine multidisciplinary teams working within nonclinical environments are yet to be understood through a human systems perspective.

Objective: This study aimed to holistically model mass COVID-19 vaccination clinics in the Region of Waterloo, Ontario, Canada, to understand the challenges centered around frontline workers and to inform clinic design and technological recommendations that can minimize the systemic inefficiencies that contribute to workplace stress.

Methods: An ethnographic approach was guided by contextual inquiry to gather data on work as done in these ad-hoc immunization settings. Observation data were clarified by speaking with clinic staff, and the research team discussed the observation data regularly throughout the data collection period. Data were analyzed by combining aspects of the contextual design framework and cognitive work analysis, and building workplace models that can identify the stress points and interconnections within mass vaccination clinic flow, developed artifacts, culture, physical layouts, and decision-making.

Results: Observations were conducted at 6 mass COVID-19 vaccination clinics over 4 weeks in 2021. The workflow model depicted challenges with maintaining situational awareness about client intake and vaccine preparation among decision-makers. The artifacts model visualized how separately developed tools for the vaccine lead and clinic lead may support cognitive tasks through data synthesis. However, their effectiveness depends on sharing accurate and timely data. The cultural model indicated that perspectives on how to effectively achieve mass immunization might impact workplace stress with changes to responsibilities. This depends on the aggressive or relaxed approach toward minimizing vaccine waste while adapting to changing policies, regulations, and vaccine scarcity. The physical model suggested that the co-location of workstations may influence decision-making coordination. Finally, the decision ladder described the decision-making steps for managing end-of-day doses, highlighting challenges with data uncertainty and ways to support expertise.

Conclusions: Modeling mass COVID-19 vaccination clinics from a human systems perspective identified 2 high-level opportunities for improving the inefficiencies within this health care delivery system. First, clinics may become more resilient to unexpected changes in client intake or vaccine preparation using strategies and artifacts that standardize data gathering and synthesis, thereby reducing uncertainties for end-of-day dose decision-making. Second, improving data sharing among staff by co-locating their workstations and implementing collaborative artifacts that support a collective understanding of the state of the

clinic may reduce system complexity by improving shared situational awareness. Future research should examine how the developed models apply to immunization settings beyond the Region of Waterloo and evaluate the impact of the recommendations on workflow coordination, stress, and decision-making.

(JMIR Hum Factors 2022;9(4):e39670) doi:10.2196/39670

KEYWORDS

cognitive work analysis; contextual design; COVID-19; decision making; health care system; pandemic; vaccination clinics; workplace stress

Introduction

Background

The concept of mass immunization was first introduced to the world in 1805, and it involved successfully vaccinating 100,000 people in Mexico for smallpox, albeit with significant logistical hurdles to bring the vaccine across the Atlantic Ocean [1]. More than 217 years later, when the global population faces its largest pandemic with COVID-19, mass immunization has never been a more critical public health activity [2]. However, during the current pandemic, mass immunization remains a multifaceted global health care challenge [3-5].

One significant logistical challenge for mass immunization is coordinating workflow in an environment where health care professionals can administer vaccines quickly and safely. The setting for such a large-scale public health activity is a unique health care context requiring significant planning and preparation among many individuals [6,7]. For example, mass vaccination clinics often combine the expertise of pharmacists, nurses, physicians, nonclinical staff, and volunteers working closely together [8]. Additionally, mass vaccination clinics are usually established in various settings, including schools, large vacant stores, city halls, shopping centers, places of worship, community centers, friendship centers, convention centers, sports arenas, and colleges or universities [6-10]. While these environments are unlikely to be designed to support vaccinating hundreds to thousands of individuals per day, implementing mass vaccination clinics in these settings is an effective way to immunize communities quickly and safely, especially during a pandemic [11].

Despite the effectiveness of COVID-19 mass vaccination clinics in supporting public health, they can be classified as complex health care delivery systems, partly resulting from vaccine brands that require ultra-low temperature storage and adequate time for thawing and mixing before administration [12-14]. They are also tightly coupled systems [15], where dependent and interconnected components, such as vaccine waste and preparation rates, can easily cause a chain reaction that impacts clinic flow [12-14], especially when a vaccine is scarce. However, the impact of this coupling is lessened when there is a surplus supply. Given the multitude of factors involved with operating a mass vaccination clinic in the era of COVID-19 [13], stress points and logistical challenges on the frontlines quickly become apparent [12,13,16].

Few scholarly articles have evaluated mass vaccination clinic design, implementation, and operation, to improve the workload of frontline staff. Some articles examining this topic have focused on developing or implementing clinic simulations and models to improve operating efficiency [17,18] or optimize the geographic placement of clinics [19]. Although the study of vaccine clinics is an emerging research area, few studies have used collaborative human factors–related methodologies to improve process inefficiencies related to clinic design [20]. While vaccine clinic culture and the interconnections among health care workers play critical roles during the evolving COVID-19 pandemic [12], to the best of the authors' knowledge, a human systems approach has not been applied to analyze the operation of mass COVID-19 vaccination clinics.

Objective

The complexities associated with multidisciplinary teams working in nontraditional large-scale immunization environments provide a critical opportunity to apply a systems approach to improve these work domains. A systems approach involving human factors modeling and analyses can identify areas and opportunities for improving team and organizational performance, as evidenced by the approach's usefulness in evaluating the interconnections and relationships within other aspects of the COVID-19 pandemic [21,22]. Therefore, the primary objective of this study was to systematically assess work as done in mass vaccination clinics, using a human factors approach [23], specifically through ethnography and a modified contextual design framework [24], substituting a tool from cognitive work analysis [25].

While primarily guided by the contextual design framework for its usefulness in driving system design [26,27], this study presents workplace models highlighting the relationships, constraints, and stress points related to mass vaccination clinic flow, developed artifacts, culture, physical layouts, and decision-making activities. Insights from this study can inform ways to minimize systemic inefficiencies for the coordinated preparation and delivery of vaccines by health care teams through technology and system design recommendations. Additionally, these results may inform public health decisions on developing or implementing tools and systems required to support frontline workers in response to the global challenge of coordinating and planning mass immunization events. Finally, this research further supports the importance of applying a human factors approach in complex health care settings and advocates for its application to understand and improve similar activities requiring quick and safe delivery of coordinated public health services to large populations.

Methods

Overview

This study used an ethnographic approach to collect data on human factors challenges faced by frontline workers operating mass COVID-19 vaccination clinics [28]. First, data collection and observations were guided by contextual inquiry at vaccination clinics in the Region of Waterloo. The collected data were subsequently organized and consolidated into the workplace categorizations within the contextual design framework: flow, artifacts, cultural, and physical [24]. Finally, the data were transformed into generalized representations using each workplace model from the framework to visualize the human system interconnections. Control task analysis was substituted from the cognitive work analysis framework, replacing the sequence model from contextual design. Instead, a decision ladder model was developed based on the skill-, rule-, and knowledge-based framework by Rasmussen [29].

Contextual Inquiry and Design

The contextual design framework is rooted in systems design and centered around workers. Hence, this methodology was chosen to provide a holistic understanding of the human factors challenges of mass COVID-19 vaccination clinics [26,30-32]. The initial step involves contextual inquiry, a participatory technique combining observations while engaging workers with their tasks. Therefore, while the researchers (RT and MT) observed each clinic, they also asked the clinic staff questions to gain expert knowledge about their roles and responsibilities, and understand how they may be influenced by other factors in the workplace [24,33].

The contextual design framework was modified in this study by substituting the sequence model with control task analysis, using the decision ladder model by Rasmussen [29]. This decision ladder model was 1 of 9 related models of naturalistic decision-making identified in 1989, including the cognitive continuum theory by Hammond and the recognition-primed decision model by Klein [34]. The presented models similarly proposed that knowledge from prior experiences was the mechanism influencing decision-making performance. While the sequence model from contextual design supports mapping the order of tasks and stages where decisions occur, the decision ladder model by Rasmussen was selected because it enables a rich understanding of the traversal among data processing and knowledge states throughout decision-making activities [25,35]. Moreover, the analysis method specifically identifies how a worker's cognition may move between nonsequential steps [36], similar to the model by Klein. We used the model by Rasmussen to capture vaccination clinic decision-making, exploiting the structure of the decision ladder to facilitate mapping heuristic pathways [36], which may be used to inform public health on improvements to operational guidance and system recommendations.

A guide was developed to record observation data (Table 1), including 4 sections aligned with the modeling frameworks [28,37]. Observed decision-making activities related to vaccine preparation and client intake were documented within all areas of the guide to inform the decision ladder modeling.

Table 1. Mass vaccination clinic observation guide.

Model (objectives)	Guiding questions
Flow (communication and coordination)	• What are the different roles that clinic staff play? • What are the responsibilities for each role? • How do staff use the clinic space to coordinate? • How do staff use artifacts to coordinate? • Who does staff give information to? In what form? • Who does the communication go to and from? • Is the communication important or an interruption? • What are the different roles that clinic staff play?
Artifacts (physical components that support the work)	• What is the structure and organization of work items used? • What information is there to use? How is it used? • What informal and formal notes do staff make? • How is information presented? • Do they need to customize something for their workspace? • How are the artifacts used to support the work?
Cultural (constraints on the work from policy/culture/values)	• What is the tone of the clinic? • What are the policies and constraints? How are they recorded? • What are the clinic staff's attitudes, feelings, and beliefs? • Do attitudes, feelings, and beliefs change over time?
Physical (physical structure of the work environment)	• What is the layout and physical location of the clinic? • Where are the tools and artifacts used? • What is the organization of the workstations? • Do staff workstations follow the clinic workflow? • Do clinic staff need to relocate often to work?

Observation Process and Ethics

During the observation period, the researchers (RT and MT) hung posters about observations taking place for a vaccine clinic research project. They also wore institution-branded clothing and name tags to identify themselves among clients and clinic staff. The researchers took notes following the observation guide, which were stored electronically on a secured server for access by the research team. Photos of artifacts and physical workspaces were taken to complement the written notes.

The research team also worked and volunteered for the Region of Waterloo, developing an emic perspective of operating a mass vaccination clinic during the initial COVID-19 immunization campaign. One researcher (CMB) volunteered within the clinics, supporting client flow with other volunteers, security staff, and clinic staff. Another researcher (KG) was the lead pharmacist on the vaccination task force for the Region of Waterloo and worked as a vaccine lead and immunizer in multiple clinics. The third researcher (MT) was involved with activities related to community outreach for vaccination in underserved populations. The fourth researcher (RT) was a volunteer and a pharmacy assistant, gaining first-hand experience tracking and drawing up the COVID-19 vaccines.

Ethics Approval

This study was reviewed and received approval from 2 research ethics committees: (1) the University of Waterloo research ethics and safety committee (#43288) and (2) the Tri-Hospital Research Ethics Committee (#2021-0735). The Region of Waterloo Public Health and Emergency Services, Grand River Hospital, and the Centre for Family Medicine Family Health Team also approved the study.

Data Analysis

The observation notes for each participating mass vaccination clinic in this study supported data source triangulation to develop consolidated contextual design models that holistically described the workplace [38]. Combining research expertise in health care systems, public health, and cognitive engineering methods in human factors, the research team reviewed and provided input to iterate on and refine the models, supporting an interdisciplinary analysis. The researchers met weekly throughout the observation period to reflect on and discuss the observation data, where the initial consolidated models were developed by 1 researcher (RT).

Results

Clinic Demographics

Between May and June 2021, the research team observed 6 mass vaccination clinics in the Region of Waterloo (Table 2). Approximately 27 hours were spent observing and interacting with the clinic staff at Clinic #1 and 8.5 hours at Clinic #2. For Clinics #3, #4, #5, and #6, the researchers observed the working environments and interacted with the staff for 10, 5, 5, and 16 hours, respectively.

Table 2. Mass vaccination clinic characteristics.

Clinic #	Environment type	Average clients/day[a]	Clinical team
1	University building	≤1000	Family health clinic
2	Vacant commercial warehouse	≤2000	Regional hospital
3	Unfinished public health building	≤1500	Public health
4	High school gymnasium[b]	≤500	Public health
5	Public health building[b]	≤500	Public health
6	Conference center[b]	≤5000	Public health

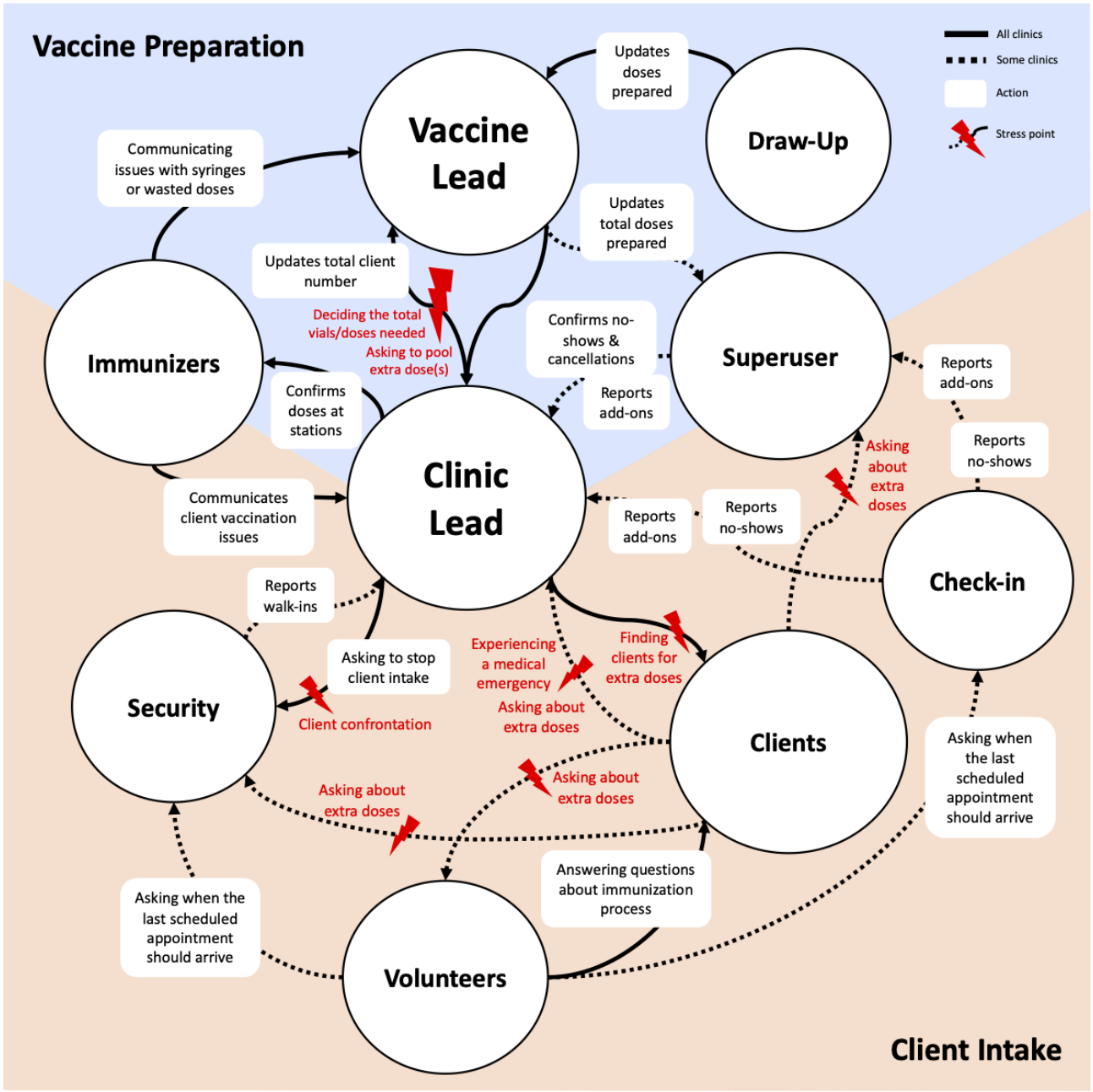
[a]Estimated value during the observation period.

[b]Primarily weekend-only clinics during the observation period.

Flow Model

Several stakeholders held multiple responsibilities contributing to workflow coordination of the mass vaccination clinics (Multimedia Appendix 1). In Figure 1, the consolidated flow model reveals stakeholder interrelations and influences between tasks and responsibilities while highlighting stress points and challenges in coordinating shared situational awareness among decision-makers. Multimedia Appendix 2 displays a more detailed version of Figure 1.

Figure 1. Consolidated flow model for mass vaccination clinics.



Information flowed among all stakeholders within the mass vaccination clinics. Salient challenges were observed between the clinic lead, superusers, and the vaccine lead, who continuously made vaccine preparation decisions to serve the constantly changing number of expected clients. However, errors in determining the accurate number of clients or prepared doses could lead to a system with higher gain, where calculation errors or unexpected cancellations result in unanticipated surplus of extra doses, causing increased stress about potential vaccine waste.

Loosening the tightly coupled vaccine preparation and client intake information was often approached through time buffering. For example, when approximately 50 expected clients remained, the clinic lead would halt client flow into the clinic. The team leads would then count the number of prepared doses remaining with immunizers and any doses still being prepared. The objective was to determine how many doses were still needed or if there might be extra doses. However, sudden unexpected changes in the number of clients and extra doses could still destabilize the system after counting, risking a surplus or shortage of doses. Regaining certainty of client intake and vaccine preparation information required a recount of available doses when client intake resumed or when initiating another pause, which could disrupt client flow.

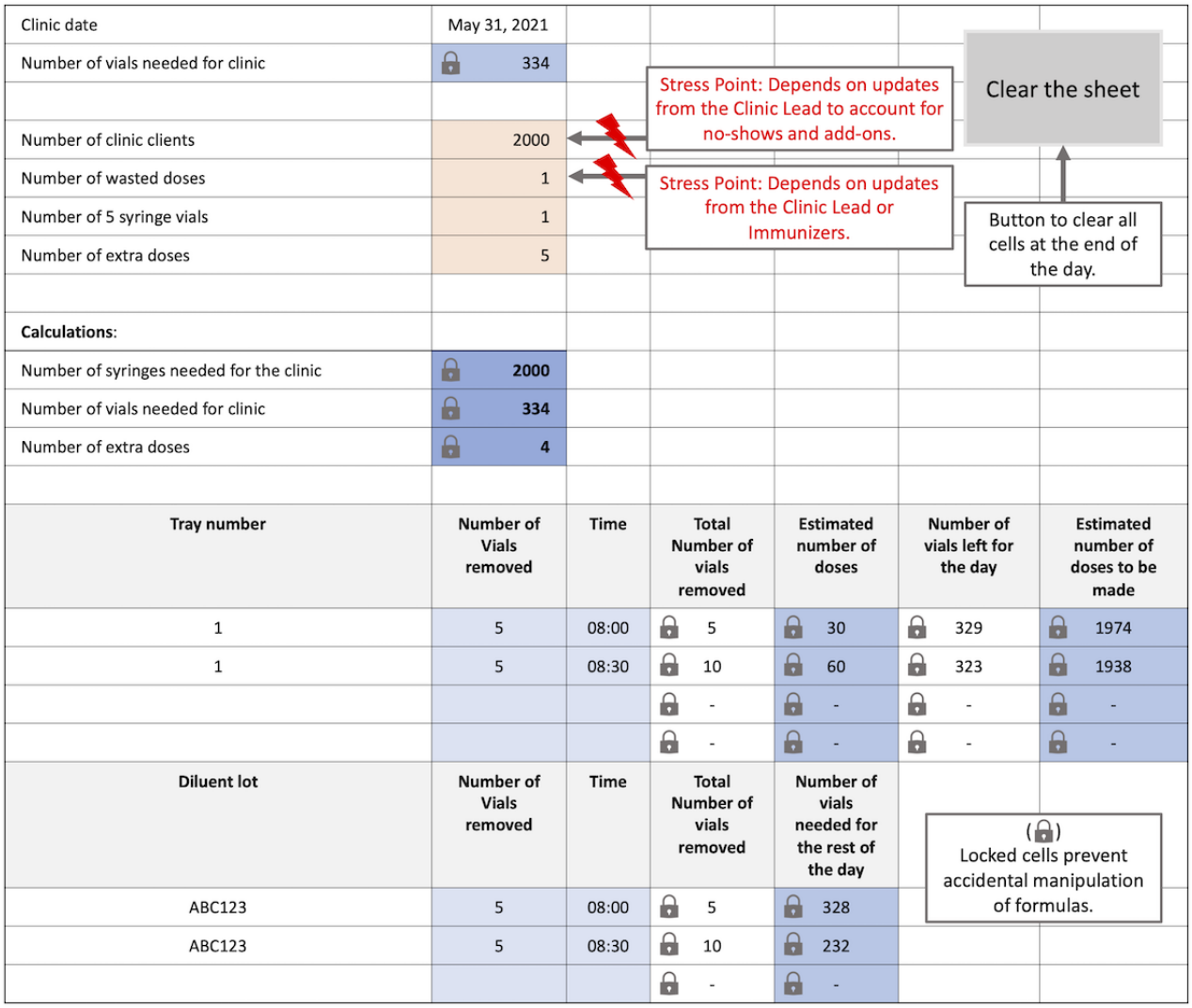
Artifacts Model

Several artifacts were observed in each mass vaccination clinic to achieve a similar intent, with a visual record of client intake and vaccine preparation changes. The different artifacts aimed to support the cognitive task of knowing how many doses were needed as early as possible but focused on tracking the details of client intake or vaccine preparation. The primary artifacts observed were categorized into 2 models: one for the vaccine lead and the other for the clinic lead.

Vaccine Lead Artifacts Model

Vaccine leads typically modified paper-based or digital artifacts to keep running totals of vaccine preparation that could be reported on public health forms at the end of the clinic day. A consolidated model combining aspects of digital and paper-based artifacts (Figure 2) is depicted as a digital spreadsheet.

Figure 2. Vaccine lead artifact model for vaccine and client tracking.



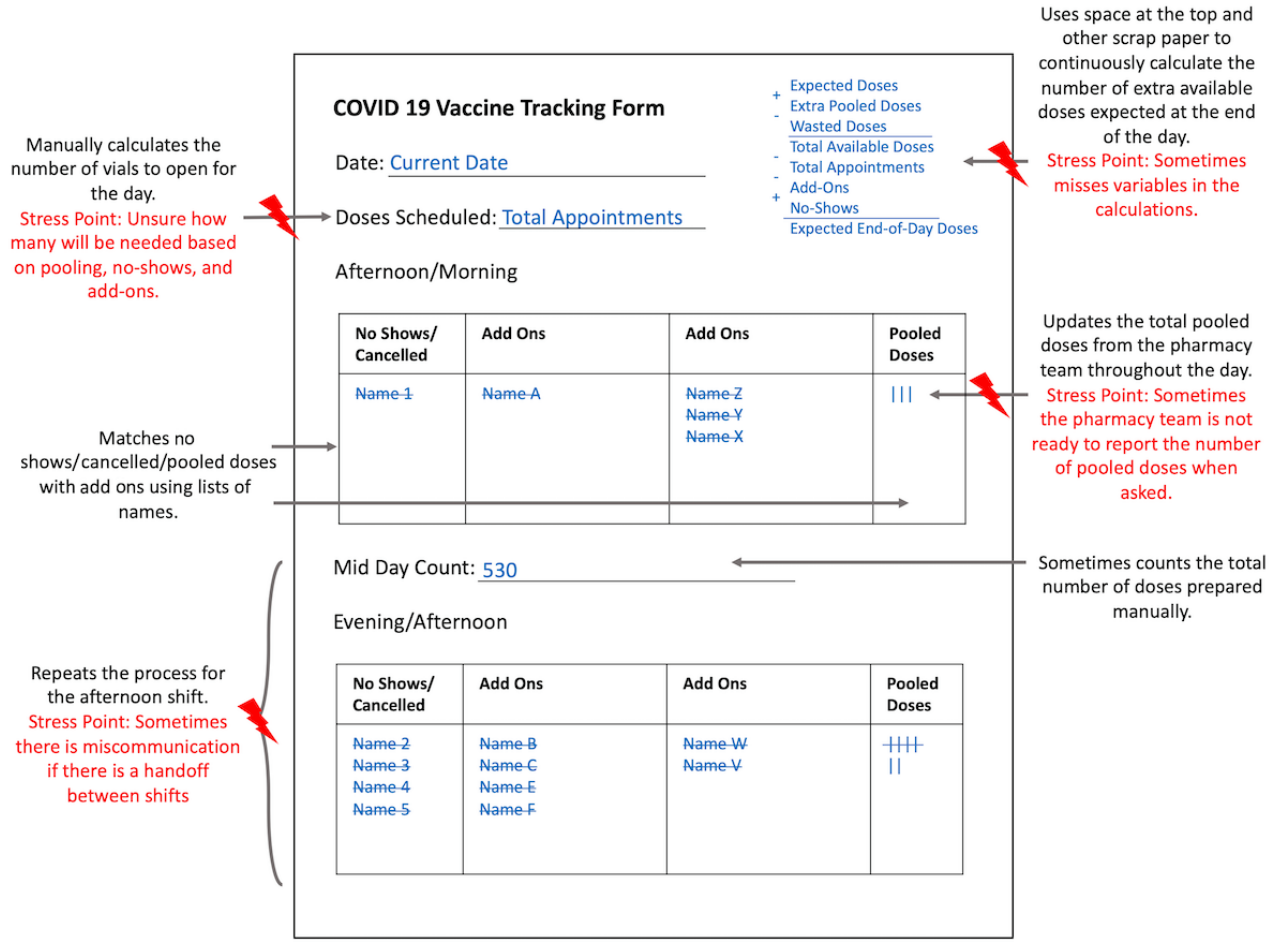
While the tracking artifacts evolved by requiring more information as the pandemic changed, both artifacts were modified differently (writing in the margins or adding new cells). The terms used in these artifacts also differed between clinics. For example, extra doses created from the residual volumes of fully prepared vials were either referred to as a "pooled dose" or a "residual dose." For clinics that provided 2 brands of vaccines, the vaccine tracking artifact would be duplicated to keep the dose tracking separated.

The initial calculations for the expected number of doses required information on the total number of clients. These data were partly obtained from the client booking website. However, the booking website could not account for client add-ons or walk-ins. Challenges resulted from not having timely updated information on client intake data from the clinic lead and the calculations from their separate tracking artifact. Unknown wasted doses at immunization stations further contributed to vaccine preparation uncertainty, limiting how effectively these tools could support decision-making if they were not accurately updated in real time.

Clinic Lead Artifacts Model

The clinic leads and superusers also developed artifacts ranging from paper-based tools to digital spreadsheets to track the number of clients expected to be immunized. In contrast to the vaccine lead artifact model, this analysis was consolidated through a paper-based representation. The paper-based model best depicts the mental models of clinic leads and superusers for client intake management (Figure 3).

Figure 3. Clinic lead artifact model for vaccine and client tracking.



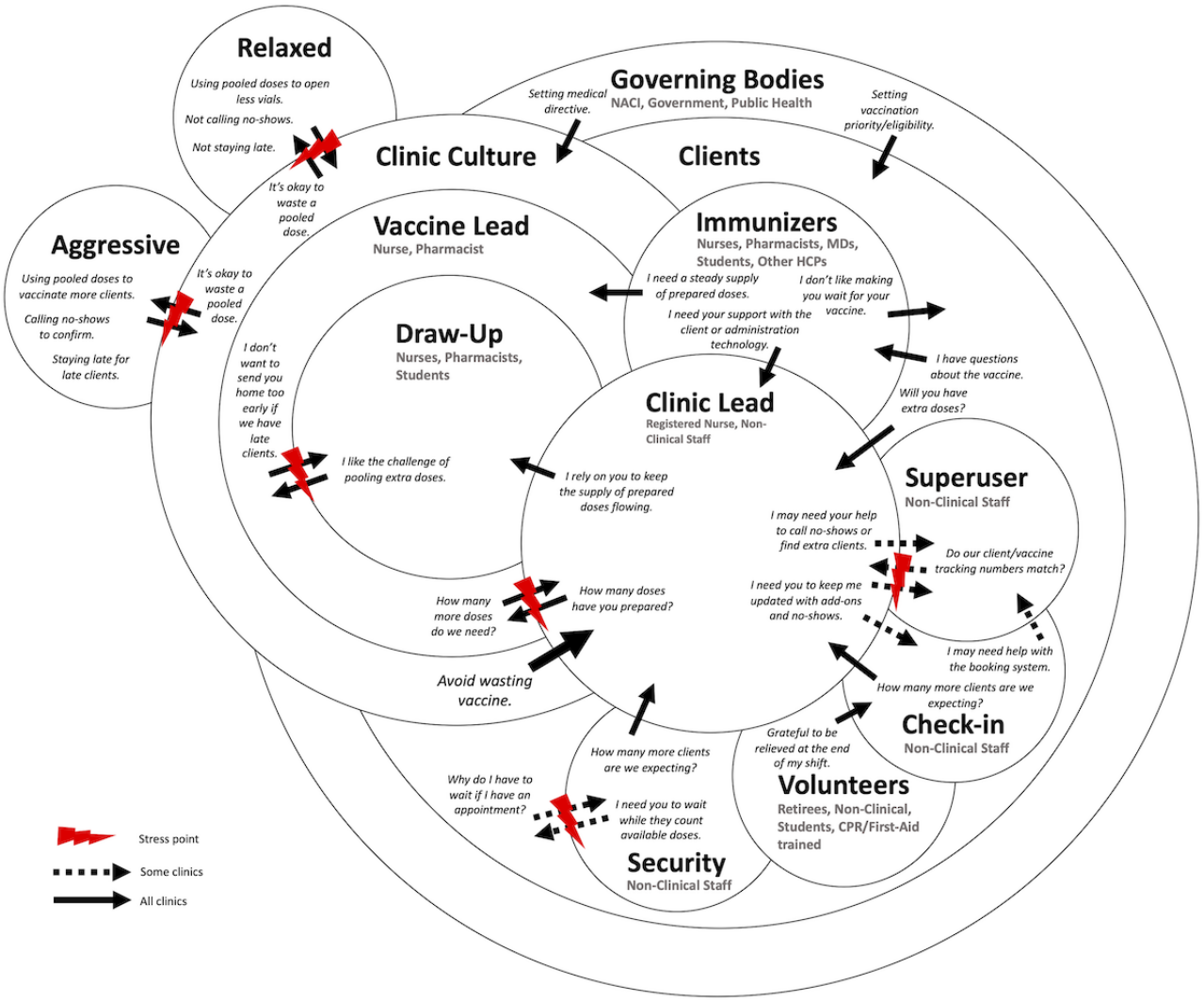
The most critical challenge was using the client intake and vaccine preparation information to support decision-making on the number of vials to open and doses to prepare. This was partly due to uncertainties in client arrival behaviors. Timely updates from the vaccine lead on the status of pooling extra doses, which varied depending on the vaccine brand, lot number, and style of syringes and needles being used (ie, varying the amount of dead volume), also contributed to uncertainty. A simpler version of the client tracking paper artifact did not include information on total vials or doses, which was used by clinic leads, superusers, and other staff who did not hold responsibilities regarding vaccine preparation decisions.

While Figure 3 is a consolidated representation of the clinic lead artifacts used to track clients and doses, versions of this tool were also implemented in digital spreadsheets at clinics working with more than one vaccine brand. The primary benefits of the digital implementation were automating calculations and providing greater granularity of tracked information. For example, while some spreadsheets categorized no-shows, deferrals/refusals, immunization ineligibilities, and duplicate appointments into 1 category, others maintained these separate categorizations. Additional data included in digitalized versions were the number of client walk-ins and add-ons from staff or volunteers being immunized; clients that requested their information be excluded from the province's immunization record website; the number of 5- and 7-dose vials; and wasted, pooled, and offsite doses brought to the clinic. A vaccine preparation rate calculation in response to the total number of immunizations over time was also included.

Cultural Model

Two workplace cultures were observed in the mass vaccination clinics: (1) the aggressive approach and (2) the relaxed approach (Figure 4). These cultures were primarily influenced by attitudes toward maximizing client throughput and managing vaccine waste, but also adapted in response to governing policies on vaccine preparation (eg, pooling residual doses), updated client eligibility, and vaccine scarcity. Client intake tracking remained the responsibility of the clinic lead in both approaches. However, the responsibility for tracking vaccine preparation and calculating the number of expected extra doses remaining at the end of the day fell on the clinic lead in the aggressive approach and the vaccine lead in the relaxed approach.

Figure 4. Consolidated cultural model for mass vaccination clinics. CPR: cardiopulmonary resuscitation; HCP: health care professional; MD: medical doctor; NACI: National Advisory Committee on Immunization.



Aggressive Approach

The aggressive approach aimed to vaccinate as many clients as possible with the amount of available vaccine. The vaccine draw-up team would create pooled doses from residual volumes shortly after preparing the expected number of doses from each vial while generally opening most of the vials planned for the number of appointments. This approach was primarily observed at smaller clinics with enough vaccines for 1 day, and increased the stress and the satisfaction of vaccinating more eligible community members. In times of vaccine scarcity, it was sometimes considered wasting vaccine if there was no attempt to pool residual volumes. It was also seen as an exciting challenge by draw-up teams, which required developing skills and new techniques to retrieve the remaining volumes of the vaccine in the vials. Clinic leads also saw it as an exciting challenge to achieve a new daily vaccination record for the clinic. They called clients who did not attend their appointments to confirm whether they would arrive later for their vaccine. However, this approach could create stress about finding eligible clients for extra doses without keeping staff, immunizers, and volunteers beyond the clinic's scheduled operating time.

Relaxed Approach

The relaxed approach aimed to vaccinate at maximum the number of scheduled clients, with the expectation of a decrease in this number throughout the day with no-shows or cancellations. This approach, used at larger clinics with more than a day's worth of stored vials, additionally used residual pooling to open fewer vials. For example, 1 less vial was opened for the Pfizer-BioNTech vaccine for every 6 pooled doses. The vaccine was still treated like it was scarce, but there was less pressure to pool residual volumes.

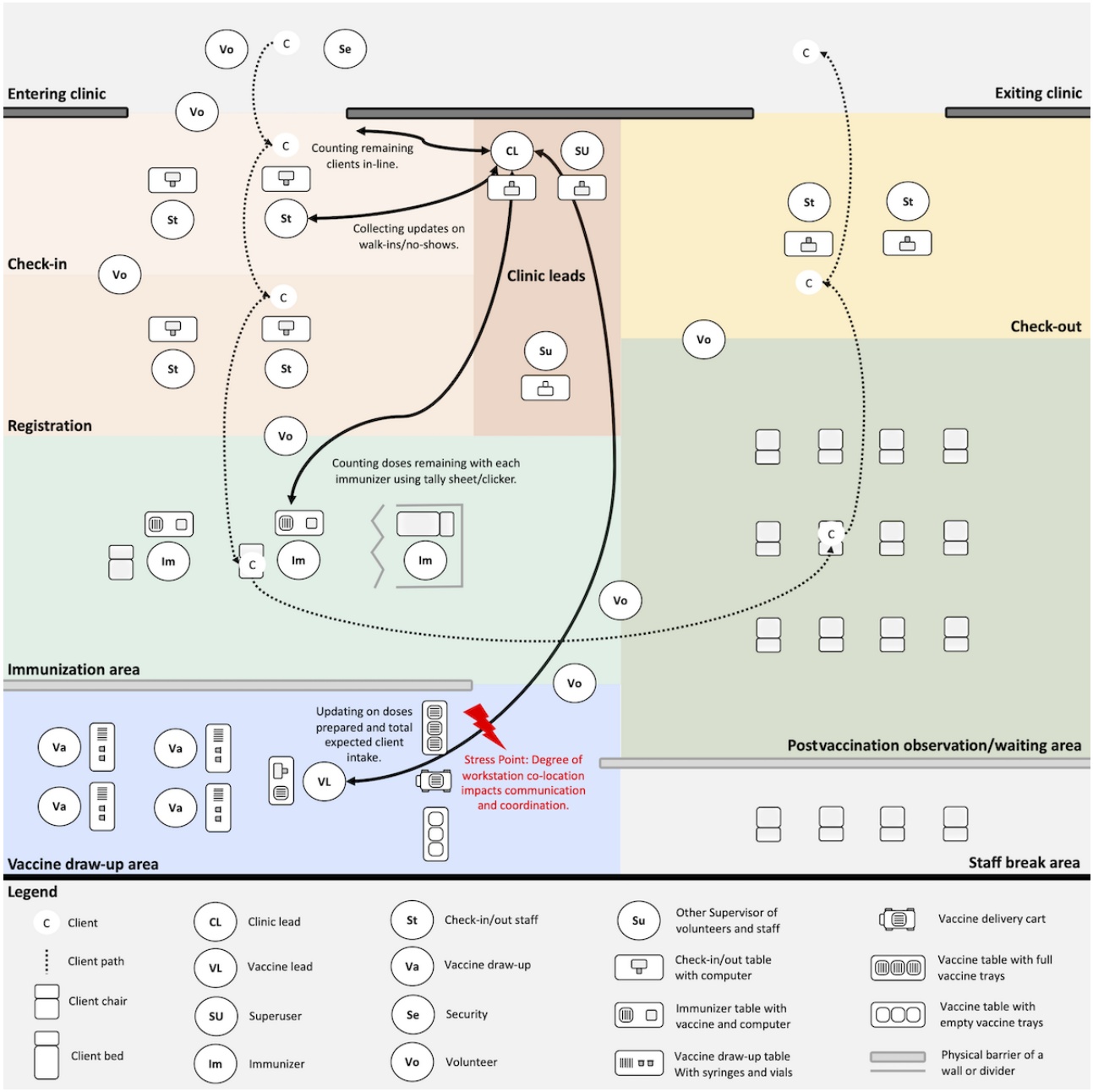
As vaccine scarcity changed, clinics would decide to pool doses within their first batch of thawed vials or during the first half of the clinic, saving the remaining residual volumes for the end of the day. If clinic leads determined that they would fall short of doses, they would inform the vaccine leads to pool residual volumes from the vials that had not been pooled yet. The remaining residual volumes that were not pooled would be discarded. In larger clinics, it was also noted that clinic leads would call clients who did not show up for their appointment. However, this task was sometimes delegated to a superuser. This approach often resulted in fewer extra doses at the end of the clinic day. However, it remained a challenge for the draw-up

team to know if they had enough residual vial volumes to meet client intake demand without having to start thawing more vials early enough to prevent delays, especially considering the expiration window.

Physical Model

The mass vaccination clinics originally consisted of the following 6 stations for clients to move between: (1) COVID-19 symptom screening, (2) checking-in for the appointment, (3) registering their personal health information, (4) receiving their vaccine, (5) sitting in the postvaccination observation and waiting area, and (6) checking-out of the clinic, where they received their vaccination receipt and, if applicable, booked their next appointment (Figure 5). Depending on the physical environment, some clinics implemented each station within a large open space or within separate rooms.

Figure 5. Consolidated physical model for mass vaccination clinics.



Each clinic's physical environment influenced where the clinic lead and the vaccine lead established their primary workstations. The larger the clinic, the more separated they were. Often, the vaccine lead worked near the immunization area, whereas the clinic lead maintained their workstation closer to the clinic entrance.

The team leads maintained frequent communication about vaccine preparation and client intake when their workstations were co-located. They were observed to make decisions quickly, and approaching each other for updates was less of an interruption of their work. It was easier to create shared awareness about the clinic's status. However, when their workstations were not co-located, it was observed that frequent communication was more challenging. The greater physical distance appeared to reduce the efficiency of communicating updates and maintaining an accurate understanding of the

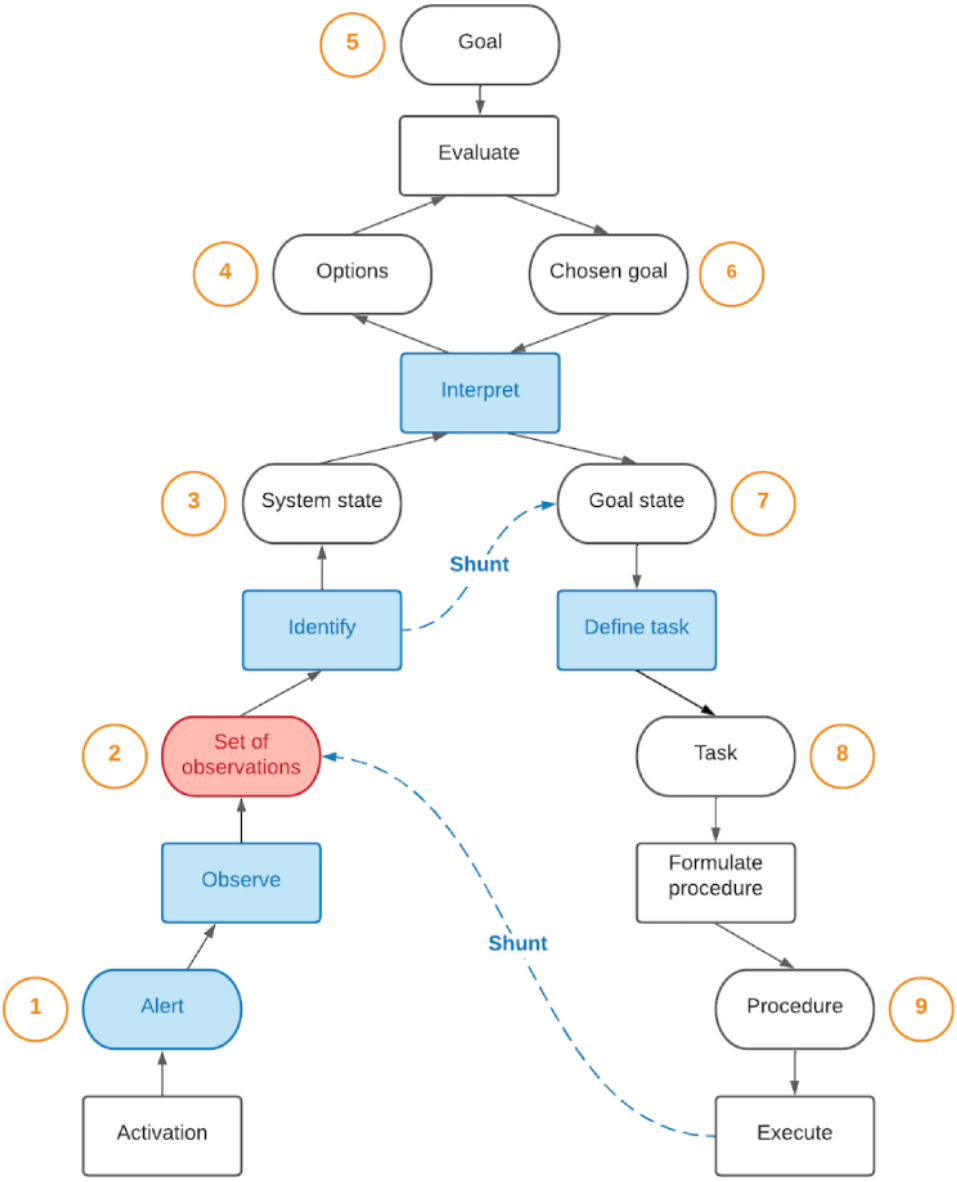
clinic's status. Strategies to overcome the physical space included writing on large whiteboards visible across the clinic and setting expectations for sharing updates at specific times.

Decision Ladder Model

The primary decision-making activity identified within mass vaccination clinics surrounded the function of avoiding vaccine waste. It was critical to understand the state of the vaccine clinic concerning the number of expected end-of-day doses, which influenced the subsequent actions the team leads would take to reduce the number of wasted doses. The clinic leads determined how many extra clients would be required for extra doses, when to start looking for additional clients to ensure they would arrive before the end of the clinic day, and where to find eligible clients. The vaccine leads determined when to stop vaccine draw-up, if they should draw up doses from new vials, and if they should draw up doses from residual vial volumes to create pooled doses.

In Figure 6, data processing activities and knowledge states in the decision ladder model that were influenced by uncertainty beyond a decision-maker's control are highlighted in red. Additionally, specific knowledge states and data processes that would benefit from support through clinic process redesign or new technologies are highlighted in blue. Shunts were present between data processing activities and the system states that expert decision-makers effectively jumped between, indicating areas within the decision-making activity of avoiding vaccine waste where introducing supports may also assist novice decision-makers.

Figure 6. Consolidated decision ladder model for mass vaccination clinics.



Step 1: Alerts

Alerts are provided by multiple sources as follows: (1) a clinic staff member, (2) a superuser, (3) a security staff member, (4) the vaccine lead, (5) the clinic lead, (6) an immunizer, (7) the booking website, and (8) the vaccine administration website. Alerts are often provided actively by the sources of the information once updates are available, or the decision-maker actively asks for new information. The alert state may be supported through technology by consolidating clinic

information within a single location or digital database that each information source can access, along with establishing processes and expectations for when information sources shall alert the decision-maker of updates.

Step 2: Set of Observations

The objective of this step is to collect and understand information on total values from individuals who are collecting data concerning the following: appointments; add-ons or walk-ins; no-shows, cancellations, deferrals, refusals, or ineligibilities; opened vials; remaining vials; pooled or residual doses; doses in preparation; doses remaining with immunizers; expected residual doses; wasted doses; and administered doses.

Uncertainty exists with some of the collected data. The booking website must accurately represent the total number of appointments, no-shows, or cancellations, as undiscovered duplicate appointments will overestimate the number of required doses. Also, if there is a lack of coordination between individuals tracking walk-ins, there is a risk that these clients could be double counted.

Similar uncertainty exists with the accuracy of the vaccine administration website representing the total number of administered doses, which assumes immunizers entered the event correctly. Additionally, if an immunizer does not vaccinate a checked-in client due to ineligibility, they need to inform someone about the extra dose. Clients who do not want their information stored within the province's online system will also not be included on the vaccine administration website and must be accounted for on a paper form.

Every event that impacts the number of required doses is essential information for the decision-maker. Therefore, this stage may also be supported by consolidating information within a single location, shared database, or collaborative interface.

Expert decision-makers may directly advance to this stage after executing tasks. After acting on a decision and updating the corresponding client intake or vaccine preparation information, they immediately understand the new state of the clinic concerning anticipated end-of-day extra doses.

Step 3: System State

At this stage, the decision-maker synthesizes the observed client intake and vaccine preparation information to determine how many available doses they expect to have at the end of the day. However, synthesizing data may be challenging if it dynamically changes. The decision-maker will often count the number of doses remaining with immunizers and the draw-up team to confirm the outcome of their calculations. Therefore, the system state may be supported by automaticity in the previous information processing stage, reducing the cognitive burden of manually performing calculations.

Step 4: Options

The decision-maker is generally faced with the following options: opening another vial, pooling residual vial volumes, finding clients for extra available doses, or sending additional clients away. The goal is to avoid significant perturbations in the state of the clinic (ie, needing significantly more doses or clients), requiring the reversal of the decision outcome if new information emerges.

Step 5: Goal

The overarching goal is to avoid vaccine waste. The potential for waste impacts overall stress with rising concerns about not having enough eligible clients to vaccinate. However, this objective is often balanced with the secondary aim of vaccinating as many eligible clients as possible, considering the potential impact on public health from contracting COVID-19 when unprotected by a vaccine.

Step 6: Chosen Goal

The decision-maker will decide how to balance maximizing vaccinations and minimizing wastage. However, they also aim to minimize extending clinic hours out of concern for burnout and client safety, and considering that extra clients are unlikely to approach the clinic for extra doses after the clinic is officially closed. The chosen approach depends on the culture of the clinic.

Step 7: Goal State

The decision-maker will determine how many additional clients to call in to get vaccinated, how many extra vials to open (or not to open), and how many extra clients must be told that there are no additional extra doses. A shunt from the *Set of Observations* stage points toward the *Goal State* due to the potential for support through automation or expertise. Calculations to determine the total available extra doses can be automated or effectively determined by an expert decision-maker, indicating whether there is a deficit or surplus in vaccines and therefore a need to draw up more doses, call in more clients, or notify clients that there are no doses available for them.

Step 8: Task

The decision-maker will decide when to notify additional clients about extra available doses, notify the draw-up team about opening extra vials, or stop opening vials, or choose when to tell clients that there are no more available doses. There is potential for supporting this stage by introducing technology that can provide real-time updates and calculations on the total available extra doses as new information emerges, given the dynamic nature of the clinic status. Real-time updates may improve decision-making efficiency.

Step 9: Procedure

Finally, the decision-maker will notify additional clients about extra available doses using public health and internal short-notice call lists, while also informing staff and volunteers within the vaccination clinic who are eligible. This process can become stressful if clients from these lists respond that they have already received their vaccine, do not answer their phone, or cannot arrive quickly. Consequently, the decision-maker may inform caregivers of clients within the vaccination clinic who are eligible for immunization. In dire situations, the decision-maker will look for extra clients outside the clinic by advertising extra doses to passers-by or within nearby commercial buildings. The final resort to avoid vaccine waste is to send extra doses to a larger clinic. Depending on the

number of expected extra clients, the decision-maker may request the draw-up team to prepare more doses by thawing a new vial (if every dose can be administered) or pool extra doses. Other actions that can be taken to avoid vaccine waste are to stop opening vials and to open fewer vials closer to the end of the clinic.

Discussion

Principal Findings

The objective of this study was to systematically evaluate *work as done* in mass vaccination clinics using the systems approach of a modified contextual design framework, highlighting the relationships and challenges related to mass vaccination clinic flow, developed artifacts, culture, physical layouts, and decision-making. The application of the contextual design framework is discussed within the dynamic nature of the COVID-19 pandemic, consolidating information across different mass immunization environments while policies and procedures were changing. We also discuss recommendations for improving the design of mass vaccination clinics and the potential for integrating digital technologies to support workflow coordination and cognitive work.

Consolidating Multiclinic Characteristics

Consolidated contextual design models are foundational to building technologies that can be better integrated into the workplace and can enhance the working environment [24]. The overall framework recognizes that introducing new technologies will inherently change the working environment; therefore, it is critical to holistically understand the working environment as it exists to provide recommendations and design solutions with a greater chance for user adoption [33]. While we modified the contextual design process to include control task analysis, the resulting decision ladder from this process enhances the understanding of cognitive processes, including differences between novice and expert decision-makers [36,39,40]. It also models how tasks and procedures are executed to achieve the desired system state [25,41].

In health care, an understanding of the contexts and processes of a work domain plays a significant role when implementing design recommendations or new technologies for health care professionals [42-44]. Successful implementation requires designing for the inherited context by matching the working culture, information needs, and work as done without introducing a significant workload [23,42,45]. It is also critical to recognize the importance of designing for humans as an adapting component in the system [23].

Only 1 descriptive report in the context of COVID-19 has provided a narrative description of vaccine clinic culture [12], pointing toward the unique challenges of designing for this health care context. In our study, the cultures of the observed mass vaccination clinics concerning role responsibility specificity, the approach for running the vaccination clinics, and the approach for handling end-of-day doses played critical roles in clinic workflow differences. Vaccine clinic cultures also indicated how workload stress might be distributed among decision-makers. Without a standardized approach to running mass vaccination clinics and individual differences between staff attitudes and beliefs across sites, this can pose a challenge for developing technological solutions that can be universally implemented to solve systemic inefficiencies.

Other nuances among mass vaccination clinics may additionally hinder the implementation of design recommendations and new technologies. This includes designing for inconsistent terminology and communicating the purposes behind recommendations when clinics approach mass immunization differently. For example, the language difference between clinics referring to extra doses as “pooled” or “residual” doses is an unsurprising inconsistency. Issues with terminology have been similarly identified in other health care contexts like medication prescribing [46]. The language inconsistencies observed in this study are likely due to the developing nature of the pandemic and different health care groups overseeing their respective clinics.

Performing contextual inquiry and design in real time has previously been successful in effectively developing health care technologies [47,48]. This study’s ethnographic approach captured the human factors necessary to understand system design within the disparate mass vaccination settings observed. The observed clinics in this study were united by the overarching goal of immunizing eligible clients and minimizing vaccine waste. Despite the cultural nuances, physical environments, and differences in staff responsibilities, the consolidated findings inform the development of generalized design recommendations, strategies, and new technologies to support these overarching objectives.

Recommendations for Improving Mass Vaccination Clinics

The developed models from this study collectively identify the need to support the limitations of information certainty (Textbox 1) and information sharing among decision-makers (Textbox 2). While considering the potential variations in the approach to mass immunization, these recommendations are expected to be applicable to improve mass immunization clinics regardless of the nuances that make them unique.

Textbox 1. Mass vaccination clinic recommendation #1.

Implement strategies and artifacts that reduce uncertainties for collecting and synthesizing client intake and vaccine preparation information required for end-of-day dose decision-making.

Textbox 2. Mass vaccination clinic recommendation #2.

Improve data sharing and its collective interpretation among staff by co-locating workstations and implementing collaborative artifacts that support shared situational awareness about the state of the clinic.

First, information uncertainty creates a high gain system (ie, where sudden changes can result in surplus vaccines) and impacts decision-making processes, especially in health care [49]. This uncertainty results in behaviors that aim to avoid decision regret and may be suboptimal for overall system performance [50]. Therefore, mass vaccination clinics should be designed to support certainty in information about client intake and vaccine preparation to reduce this associated stress. One potential strategy includes always using an appointment-based system to increase information certainty for how many clients are expected throughout the day and the number of doses to prepare. As previously identified, appointment-based systems double as a strategy to prevent bottlenecks in clinic flow by controlling the maximum client intake volume [51]. A layer of certainty to client intake can also include assigning staff to call no-shows and confirm their arrival, as identified by the aggressive cultural approach to operating a clinic. From the vaccine lead perspective, strategies to reduce uncertainty include the early identification of the expected number of doses per multidose vial, using the style of syringes and needles provided for the draw-up, while also accounting for potential equipment shortages that may influence the total amount of expected doses (ie, changes in syringe dead volume). From the relaxed approach, reducing the number of new vials required for every vial's worth of pooled residual doses may decrease the risk of a surplus or scarcity of doses in a future state of the clinic.

Digital technologies can also reduce information uncertainty in mass vaccination clinics to support overall system resilience while potentially being broadly adopted by frontline workers, as seen in prior mass immunization contexts [52]. Information certainty can be improved within data processing tasks by reducing human error, which can support the accuracy of the calculations required to determine the numbers of anticipated end-of-day extra doses, remaining clients, and remaining doses through automation in real time. Automating the calculations needed to operate a mass vaccination clinic may standardize end-of-day dose decision-making regardless of the clinic culture. Subsequently, if digital tools can be widely adopted, this may standardize data processing activities across more than one clinic.

Second, information sharing among data source stakeholders is essential for successfully operating a mass vaccination clinic. Shared information supports decision-making tasks through a collective understanding of the dynamic state of the clinic and how it might change in the future (ie, the degree of vaccine excess or vaccine shortage). Shared interprofessional decision-making is fundamental in health care [49]. In the context of mass vaccination clinics, this process occurs between the clinic lead and the vaccine lead. It is critically essential to have the information and the skills that enable shared knowledge of the current and future states of the clinic when changes in client intake and vaccine preparation occur.

As observed in this study, increased co-location is one way to support shared awareness between staff managing client intake and vaccine preparation information, and decision-making. Additionally, establishing communication expectations for when and how to update information between staff members was observed to support situational awareness when followed (eg, updating on the hour using a standard template). While establishing clinics in large open spaces has been previously shown to improve workflow efficiencies with clients due to greater flexibility in workstation placement [20], clinics operating out of such environments should consider primary workstation placements that reduce the physical separation of decision-makers and the data they need.

However, the rigidity of the physical environment may prevent co-locating workstations, and decision-makers may be moving throughout the clinic, creating a cognitive burden to gather and synthesize information partly in isolation. One solution is to develop and implement collaborative artifacts that support situational awareness through real-time shared databases, computerized applications, or large displays (eg, a large whiteboard or poster that can be viewed across the clinic). This would reduce the need to *physically* gather dynamic data, similar to the workflows within intensive care units [53]. In vaccine clinics, implementing artifacts that automatically synthesize information in a standard way through a shared interface would further bridge the artifact silos that currently exist. If intuitively displayed, this would support a shared understanding of the state of the clinic when co-locating workstations is infeasible.

Strengths and Limitations

This is the first study to apply a human systems approach to mass vaccination clinics during the COVID-19 pandemic with an interdisciplinary research team. However, while the comprehensive data collected from this study included the Region of Waterloo in Ontario, Canada, the results have not been validated in other regions or within resource-limited settings. Although the researchers confirmed the raw observation data with staff, only the research team reviewed the resulting models.

While this study captures important high-level insights during a critical period of the COVID-19 pandemic in the Region of Waterloo, we recognize that it does not entirely capture the complexities of all mass vaccination environments or precisely replicate real-world scenarios throughout the pandemic. Therefore, further research should examine the inclusion of additional complexities in the developed models and their application beyond mass vaccination clinics in the Region of Waterloo. Future work can also evaluate the recommendations on workflow coordination, stress, and decision-making. Finally, microergonomic analyses aimed at improving specific processes in mass vaccination clinics could support the optimal design of public health documentation and workstations for vaccine draw-up.

Conclusions

This study provides a holistic representation of the working environments in mass vaccination clinics within the Region of Waterloo, highlighting work as done and human factors challenges in the following human systems models: flow, artifacts, cultural, physical, and decision-making. A systems analysis using human factors methodologies has provided a critically important lens on mass vaccination clinics in the context of COVID-19. The collaboration between human factors

researchers and health care professionals in this study has also shown essential advantages for effectively understanding and modeling this complex health care environment and providing important recommendations to improve this complex tightly coupled system. While this research provides a fundamental understanding of mass vaccination clinics centered around frontline workers, further collaborative research that bridges human factors and health care professionals can advance the scope of knowledge on improving this evolving global public health activity.

Acknowledgments

We appreciate the willingness of the staff at the mass vaccination clinics in the Region of Waterloo to allow us to observe their work throughout the ongoing COVID-19 pandemic. This research was supported through funding from the Region of Waterloo and a Natural Sciences and Engineering Research Council Collaborative Research and Development Grant. However, they were not involved in study design, data collection, or formal analysis. RT is funded by the Engineering Excellence Doctoral Fellowship from the University of Waterloo.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Observed workflow-related tasks and responsibilities in mass vaccination clinics.
[PDF File (Adobe PDF File), 136 KB - humanfactors_v9i4e39670_app1.pdf]

Multimedia Appendix 2
A detailed version of the consolidated flow model for mass vaccination clinics.
[PDF File (Adobe PDF File), 117 KB - humanfactors_v9i4e39670_app2.pdf]

References

1. Mark C, Rigau-Pérez JG. The world's first immunization campaign: the Spanish Smallpox Vaccine Expedition, 1803-1813. Bull Hist Med 2009;83(1):63-94 [FREE Full text] [doi: 10.1353/bhm.0.0173] [Medline: 19329842]
2. Goralnick E, Kaufmann C, Gawande AA. Mass-Vaccination Sites — An Essential Innovation to Curb the Covid-19 Pandemic. N Engl J Med 2021 May 06;384(18):e67. [doi: 10.1056/nejmp2102535]
3. Rigby MJ. After the European Global Vaccination Summit-The need for practical policies to boost vaccination delivery to children in Europe. Int J Health Plann Manage 2021 Nov 13;36(6):1998-2010. [doi: 10.1002/hpm.3282] [Medline: 34254365]
4. Dhama K, Sharun K, Tiwari R, Dhawan M, Emran TB, Rabaan AA, et al. COVID-19 vaccine hesitancy - reasons and solutions to achieve a successful global vaccination campaign to tackle the ongoing pandemic. Hum Vaccin Immunother 2021 Oct 03;17(10):3495-3499 [FREE Full text] [doi: 10.1080/21645515.2021.1926183] [Medline: 34191680]
5. Sayegh MH, Merhi A, Naal H, Saleh S. Global vaccination strategy to eradicate COVID-19: Beyond science. J Glob Health 2021 Nov 13;11:03113 [FREE Full text] [doi: 10.7189/jogh.11.03113] [Medline: 34858592]
6. Park K, Cartmill R, Johnson-Gordon B, Landes M, Malik K, Sinnott J, et al. Preparing for a School-Located COVID-19 Vaccination Clinic. NASN School Nurse 2021 Feb 22;36(3):156-163. [doi: 10.1177/1942602x21991643]
7. Planning guidance for immunization clinics for COVID-19 vaccines. Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/guidance-documents/planning-immunization-clinics-covid-19-vaccines.html#shr-pg0 [accessed 2022-10-26]
8. Cummings G, Ruff E, Guthrie S, Hoffmaster M, Leitch L, King J. Successful use of volunteers to conduct school-located mass influenza vaccination clinics. Pediatrics 2012 Mar;129 Suppl 2(Supplement_2):S88-S95. [doi: 10.1542/peds.2011-0737H] [Medline: 22383487]
9. Sreenivasan MV, He HH, Park SY. Administration time between seasonal live-attenuated influenza vaccine and trivalent influenza vaccine during the "Stop Flu at School" Campaign--Hawaii, 2009. Public Health Rep 2014 May 01;129(3):229-236 [FREE Full text] [doi: 10.1177/003335491412900304] [Medline: 24791020]
10. Fontanesi J, Jue-Leong S. Logistical and fiscal sustainability of a school-based, pharmacist-administered influenza vaccination program. Journal of the American Pharmacists Association 2012 Sep;52(5):e74-e79. [doi: 10.1331/japha.2012.10113]
11. Fontanesi J, Hill L, Olson R, Bennett NM, Kopald D. Mass vaccination clinics versus appointments. J Med Pract Manage 2006;21(5):288-294. [Medline: 16711096]
12. Ezeude G, Glover K, Nieves Santiago AA, Repella E, Tang E. Roadblocks and successes in preparing COVID-19 vaccination clinics: Perspectives from pharmacy residents. Am J Health Syst Pharm 2022 Jan 05;79(2):102-109 [FREE Full text] [doi: 10.1093/ajhp/zxab347] [Medline: 34487141]

13. Li Q, Norman S, Guthrie P, Gachoya J, Sebakijje J, Leftridge Z, et al. Strategies used to meet the challenges of mass COVID-19 vaccination by the pharmacy department in a large academic medical center. Am J Health Syst Pharm 2021 Sep 07;78(18):1724-1731 [FREE Full text] [doi: 10.1093/ajhp/zxab184] [Medline: 33895789]
14. Uddin MN, Roni MA. Challenges of Storage and Stability of mRNA-Based COVID-19 Vaccines. Vaccines (Basel) 2021 Sep 17;9(9):1033 [FREE Full text] [doi: 10.3390/vaccines9091033] [Medline: 34579270]
15. Perrow C. Normal Accidents: Living with High Risk Technologies. Princeton, New Jersey: Princeton University Press; 1999.
16. Andrade J, Slaby M, DeAngelis J, Connors J, Truong J, Ciaramella C, et al. Implementation of a pharmacist-led COVID-19 vaccination clinic at a community teaching hospital. Am J Health Syst Pharm 2021 Jun 07;78(12):1038-1042 [FREE Full text] [doi: 10.1093/ajhp/zxab135] [Medline: 33772261]
17. Stone K, Enriquez B, Keilman A, German K, Straight A, Demmert KR, et al. Using a Systems-Focused Simulation to Rapidly Prepare for Large Volume COVID-19 Vaccination. Sim Healthcare 2021 Mar 30;16(6):420-421. [doi: 10.1097/sih.0000000000000573]
18. Carr C, Durrheim D, Eastwood K, Massey P, Jaggers D, Caelli M, et al. Australia's first pandemic influenza mass vaccination clinic exercise: Hunter New England Area Health Service, NSW, Australia. Australian Journal of Emergency Management 2011;26(1):47-53 [FREE Full text]
19. Çakır E, Taş MA, Ulukan Z. Neutrosophic Fuzzy Weighted Saving Heuristic for COVID-19 Vaccination. 2021 Presented at: Systems and Information Engineering Design Symposium (SIEDS); April 29-30, 2021; Charlottesville, VA, USA. [doi: 10.1109/SIEDS52267.2021.9483794]
20. Ha C, Taylor C, Modi JR. Mass Vaccinations at the United States Naval Academy. Health Secur 2016 Dec;14(6):382-388. [doi: 10.1089/hs.2016.0030] [Medline: 27830937]
21. Salmon PM, Stevens N, McLean S, Hulme A, Read GJM. Human Factors and Ergonomics and the management of existential threats: A work domain analysis of a COVID-19 return from lockdown restrictions system. Hum Factors Ergon Manuf 2021 Jul 06;31(4):412-424 [FREE Full text] [doi: 10.1002/hfm.20897] [Medline: 34230798]
22. Plant KL, Fendley ME. Human factors and ergonomics and the response to COVID‐19. Hum Factors Man 2021 Jun 11;31(4):329-332. [doi: 10.1002/hfm.20923]
23. Hollnagel E, Wears R, Braithwaite J. From Safety-I to Safety-II: A White Paper. NHS. 2015. URL: https://www.england.nhs.uk/signuptosafety/wp-content/uploads/sites/16/2015/10/safety-1-safety-2-whte-papr.pdf [accessed 2022-10-26]
24. Holtzblatt K, Jones S. Contextual Inquiry: A Participatory Technique for System Design. In: Schuler D, Namioka A, editors. Participatory Design: Principles and Practices. Boca Raton, Florida: CRC Press; 1993.
25. Vicente K. Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Boca Raton, Florida: CRC Press; 1999.
26. Wixon D, Holtzblatt K, Knox S. Contextual design: an emergent view of system design. In: CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 1-5, 1990; Seattle, WA, USA p. 329-336. [doi: 10.1145/97243.97304]
27. Holtzblatt K, Beyer H. Contextual design: using customer work models to drive systems design. In: CHI '96: Conference Companion on Human Factors in Computing Systems. 1996 Presented at: Conference Companion on Human Factors in Computing Systems; April 13-18, 1996; Vancouver, British Columbia, Canada p. 373-374. [doi: 10.1145/257089.257379]
28. Hughes J, O'Brien J, Rodden T, Rouncefield M, Sommerville I. Presenting ethnography in the requirements process. In: Proceedings of 1995 IEEE International Symposium on Requirements Engineering (RE'95). 1995 Presented at: IEEE International Symposium on Requirements Engineering (RE'95); March 27-29, 1995; York, UK p. 27-34. [doi: 10.1109/isre.1995.512543]
29. Rasmussen J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Syst., Man, Cybern 1983 May;SMC-13(3):257-266. [doi: 10.1109/TSMC.1983.6313160]
30. Norman D, Draper S. User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ: L. Erlbaum Associates Inc; 1986.
31. Hefley WE, Buie EA, Lynch GF, Muller MJ, Hoecker DG, Carter J, et al. Integrating Human Factors with Software Engineering Practices. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2016 Nov 05;38(4):315-319. [doi: 10.1177/154193129403800427]
32. Schuler D, Namioka A. Participatory Design: Perspectives on Systems Design. Hillsdale, NJ: L. Erlbaum Associates Inc; 1993.
33. Beyer H, Holtzblatt K. Contextual Design: Defining Customer-Centered Systems. San Francisco, CA: Morgan Kaufmann Publishers Inc; 1997.
34. Klein G. Naturalistic decision making. Hum Factors 2008 Jun;50(3):456-460. [doi: 10.1518/001872008X288385] [Medline: 18689053]
35. Naikar N, Moylan A, Pearce B. Analysing activity in complex systems with cognitive work analysis: concepts, guidelines and case study for control task analysis. Theoretical Issues in Ergonomics Science 2006 Jul;7(4):371-394. [doi: 10.1080/14639220500098821]
36. Bisantz A, Burns C. Applications of Cognitive Work Analysis. Boca Raton, Florida: CRC Press; 2008.

37. Holtzblatt K. Customer-centered design for mobile applications. Pers Ubiquit Comput 2005 May 11;9(4):227-237. [doi: 10.1007/s00779-004-0324-5]
38. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
39. Rasmussen J. The human as a system component. In: Smith H, Green TRG, editors. Human Interaction With Computers. London, UK: Academic Press; 1980.
40. Rasmussen J. The human data processor as a system component. Bits and pieces of a model. Risø National Laboratory. 1974. URL: https://backend.orbit.dtu.dk/ws/portalfiles/portal/56551783/ris_m_1722.pdf [accessed 2022-10-26]
41. Burns CM, Vicente KJ. Model-Based Approaches for Analyzing Cognitive Work: A Comparison of Abstraction Hierarchy, Multilevel Flow Modeling, and Decision Ladder Modeling. International Journal of Cognitive Ergonomics 2001 Sep;5(3):357-366. [doi: 10.1207/s15327566ijce0503_13]
42. Boddy D, King G, Clark JS, Heaney D, Mair F. The influence of context and process when implementing e-health. BMC Med Inform Decis Mak 2009 Jan 30;9:9 [FREE Full text] [doi: 10.1186/1472-6947-9-9] [Medline: 19183479]
43. Kushniruk A, Nohr C, Jensen S, Borycki EM. From Usability Testing to Clinical Simulations: Bringing Context into the Design and Evaluation of Usable and Safe Health Information Technologies. Yearb Med Inform 2018 Mar 05;22(01):78-85. [doi: 10.1055/s-0038-1638836]
44. Patel VL, Kushniruk AW. Interface design for health care environments: the role of cognitive science. Proc AMIA Symp 1998:29-37 [FREE Full text] [Medline: 9929179]
45. Pavel M, Jimison HB, Wactlar HD, Hayes TL, Barkis W, Skapik J, et al. The role of technology and engineering models in transforming healthcare. IEEE Rev Biomed Eng 2013;6:156-177. [doi: 10.1109/RBME.2012.2222636] [Medline: 23549108]
46. Mercer K, Carter C, Burns C, Tennant R, Guirguis L, Grindrod K. Including the Reason for Use on Prescriptions Sent to Pharmacists: Scoping Review. JMIR Hum Factors 2021 Nov 25;8(4):e22325 [FREE Full text] [doi: 10.2196/22325] [Medline: 34842545]
47. Miller A, Koola JD, Matheny ME, Ducom JH, Slagle JM, Groessl EJ, et al. Application of contextual design methods to inform targeted clinical decision support interventions in sub-specialty care environments. Int J Med Inform 2018 Sep;117:55-65. [doi: 10.1016/j.ijmedinf.2018.05.005] [Medline: 30032965]
48. Jacobsohn GC, Leaf M, Liao F, Maru AP, Engstrom CJ, Salwei ME, et al. Collaborative design and implementation of a clinical decision support system for automated fall-risk identification and referrals in emergency departments. Healthc (Amst) 2022 Mar;10(1):100598. [doi: 10.1016/j.hjdsi.2021.100598] [Medline: 34923354]
49. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). Social Science & Medicine 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3]
50. Bell DE. Regret in Decision Making under Uncertainty. Operations Research 1982 Oct;30(5):961-981. [doi: 10.1287/opre.30.5.961]
51. Bourgeois N, Franke C, O'Connor SA, Shaw H, Hum S, Dunn S. Preparing for an influenza pandemic: model of an immunization clinic in an urban family practice. Can Fam Physician 2011 Oct;57(10):e367-e372 [FREE Full text] [Medline: 21998244]
52. Foisy J, Quach S, Heidebrecht CL, Pereira JA, Quan SD, Guay M, Public Health Agency of Canada/Canadian Institutes of Health Research Influenza Research Network (PCIRN) Vaccine Coverage Theme Group. Perceptions of frontline staff regarding data collection methodologies used during the 2009 A H1N1 influenza immunization campaign in Canada. BMC Public Health 2010 Dec 30;10(1):796 [FREE Full text] [doi: 10.1186/1471-2458-10-796] [Medline: 21831331]
53. Flohr L, Beaudry S, Johnson KT, West N, Burns CM, Ansermino JM, et al. Clinician-Driven Design of VitalPAD–An Intelligent Monitoring and Communication Device to Improve Patient Safety in the Intensive Care Unit. IEEE J. Transl. Eng. Health Med 2018;6:1-14. [doi: 10.1109/jtehm.2018.2812162]

Edited by A Kushniruk; submitted 07.06.22; peer-reviewed by D Tuden, P Tremoulet; comments to author 02.08.22; revised version received 15.09.22; accepted 06.10.22; published 10.11.22.

Please cite as:
Tennant R, Tetui M, Grindrod K, Burns CM
Understanding Human Factors Challenges on the Front Lines of Mass COVID-19 Vaccination Clinics: Human Systems Modeling Study
JMIR Hum Factors 2022;9(4):e39670
URL: https://humanfactors.jmir.org/2022/4/e39670
doi:10.2196/39670
PMID:36219839

Original Paper

Reinterpretation of Health Information in the Context of an Emerging Infectious Disease: A Digital Focus Group Study

Rayner Kay Jin Tan[1,2,3], PhD; Jane Mingjie Lim[2], PhD; Pearlyn Hui Min Neo[2], MA; Suan Ee Ong[2,4], PhD

[1]University of North Carolina Project-China, Guangzhou, China

[2]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[3]Dermatology Hospital of Southern Medical University, Guangzhou, China

[4]Research for Impact, Singapore, Singapore

Corresponding Author:
Rayner Kay Jin Tan, PhD
University of North Carolina Project-China
2 Lujing Road
Guangzhou, 510095
China
Phone: 65 91878576
Email: rayner.tan@nus.edu.sg

Abstract

Background: Misinformation related to the COVID-19 pandemic has accelerated global public concern and panic. The glut of information, or "infodemic," has caused concern for authorities due to its negative impacts on COVID-19 prevention and control, spurring calls for a greater scholarly focus on health literacy during the pandemic. Nevertheless, few studies have sought to qualitatively examine how individuals interpreted and assimilated health information at the initial wave of COVID-19 restrictions.

Objective: We developed this qualitative study adopting chat-based focus group discussions to investigate how individuals interpreted COVID-19 health information during the first wave of COVID-19 restrictions.

Methods: We conducted a qualitative study in Singapore to investigate how individuals perceive and interpret information that they receive on COVID-19. Data were generated through online focus group discussions conducted on the mobile messaging smartphone app WhatsApp. From March 28 to April 13, 2020, we held eight WhatsApp-based focus groups (N=60) with participants stratified by age groups, namely 21-30 years, 31-40 years, 41-50 years, and 51 years and above. Data were thematically analyzed.

Results: A total of four types of COVID-19 health information were generated from the thematic analysis, labeled as formal health information, informal health information, suspicious health information, and fake health information, respectively. How participants interpreted these categories of information depended largely on the perceived trustworthiness of the information source as well as the perceived veracity of information. Both factors were instrumental in determining individuals' perceptions, and their subsequent treatment and assimilation of COVID-19–related information.

Conclusions: Both perceived trustworthiness of the information source and perceived veracity of information were instrumental concepts in determining one's perception, and thus subsequent treatment and assimilation of such information for one's knowledge of COVID-19 or the onward propagation to their social networks. These findings have implications for how policymakers and health authorities communicate with the public and deal with fake health information in the context of COVID-19.

(JMIR Hum Factors 2022;9(4):e39312) doi:10.2196/39312

KEYWORDS

health communication; infodemic; SARS-CoV-2; coronavirus; Singapore; WhatsApp; COVID-19; health information; misinformation; mobile health; smartphone; information quality; online health information

Introduction

Background

The COVID-19 pandemic was declared a public health emergency of international concern by the World Health Organization (WHO) in January 2020 [1] and was subsequently declared a pandemic in March 2020 [2]. As of April 2022, globally, over 500 million people have been infected with the virus that causes COVID-19 and more than 6 million people have died from the disease [3]. Besides its impact on morbidity and mortality, COVID-19 has deeply impacted economies, health systems, and social lives globally [4].

The rapid spread of COVID-19 and its variants Delta and Omicron led to an urgent need for reliable, evidence-based, trustworthy, and updated health information that could help individuals to inform decision-making around COVID-19 prevention and management. Early dissemination of COVID-19 prevention guidelines by global and national authorities was accompanied by a concomitant rise of misinformation or fake news [5], termed an "infodemic" by the WHO Director-General Tedros Adhanom Ghebreyesus. The infodemic has caused concern for authorities due to its negative impacts on COVID-19 prevention and control [6], spurring calls for a greater scholarly focus on health literacy during the pandemic [7-9].

The infodemic has resulted in challenges in efficiently and trustworthily conveying reliable, rigorous COVID-19 information to the public. For example, a nationally representative online survey in Germany found that 47.8% of participants had trouble assessing if media information on COVID-19 could be trusted [10]. The infodemic has simultaneously facilitated the spread of conspiracy theories on COVID-19, including its origins and vaccines [11-13]. Studies also highlight how those with low levels of health literacy and those residing in low- to middle-income countries are at the greatest risk of succumbing to false or misleading pandemic-related information [10,14].

Past studies have attempted to determine the relationship between general literacy and COVID-19–related health literacy [15-17] or nuance the nature and types of COVID-19 information [18-20]. However, there has been minimal research to explore the factors that affect how individuals negotiate, construe, and interpret information in an infodemic. An understanding of such human factors, specifically how individuals interpret health information during emerging infectious disease contexts, is important for the widespread assimilation and integration of health information technologies in our ongoing management of the current pandemic and future pandemics.

Use of Mobile Chat Apps for Focus Group Discussions

The use of focus group discussions (FGDs) for social science and health research, especially in reference to its methodological rigor for qualitative research, has been widely researched [21,22]. Such inquiries have extended to online FGDs, with their origins extending to the late 1990s with the use of emails and message boards [23], and subsequently with virtual discussion rooms and teleconferencing software [24,25].

Studies have found that online FGDs have important benefits relative to in-person groups. While responses were typically less detail-rich, researchers found that they were more immediate. Past studies also argued that online FGDs lack contextual clues that may inform perceived power differences between participants, which may facilitate the sharing of sensitive information and disagreements among participants [26,27]. Logistically, online groups can be more inclusive than traditional in-person FGDs, reducing physical access–related barriers to participation and allowing for participants to share diverse types of media throughout the course of the FGDs [28].

We also recognize the drawbacks of online FGDs compared to in-person groups. Online FGDs tend to generate lower word counts, shorter responses, and provide less detail or richness [25,26,29]. They also tend to have fewer group interactions and lower responsiveness to facilitators' questions and probes [25,27,29].

Recent scholarship has highlighted the usefulness of mobile chat platforms such as WhatsApp in eliciting and generating qualitative data. Chen and Neo's [30] study comparing indicators of data depth and breadth between WhatsApp-based and in-person FGDs in Singapore found that while data richness and detail in the WhatsApp FGDs did not match that of in-person groups, younger and more digitally savvy participants generated well-elaborated responses and were interactive within the chat groups. Findings from Colom's [31] study in Western Kenya employing digital ethnographic methods, including the use of WhatsApp FGDs, also showed that the use of WhatsApp provided a high level of ecological validity [31].

The COVID-19 Pandemic in Singapore

The first case of COVID-19 was reported in Singapore on January 23, 2020, prompting national authorities to implement a series of movement control measures to curb disease spread. One of the first steps taken involved the closure of entertainment establishments in late February 2020 following the change in Singapore's Disease Outbreak Response System Condition (DORSCON) color code from yellow to orange. This signaled an official recognition of COVID-19's severity and infectiousness, and a recognition that the disease had arrived in Singapore. A "circuit-breaker" period was implemented from April 7 to June 1, 2020, characterized by strict movement control measures aimed at curbing the growing incidence of COVID-19 cases in the community, and a "break the circuit" of transmission [32] period, including the closure of all nonessential services and a mask mandate. Since June 2020, Singapore has gradually eased these measures in phases; phase one took place between June 2 and June 18, 2020, which involved the progressive resumption of select businesses and activities.

Study Rationale and Objective

Against the backdrop of information-related concern and policy responses in the early stages of the COVID-19 pandemic in Singapore, we developed this study to investigate how individuals interpreted COVID-19 health information during the first wave of COVID-19 restrictions.

Methods

Participants and Data Generation

We conducted a series of eight FGDs on WhatsApp between March 28 and April 13, 2020. Further details on how the FGDs were conducted have been published elsewhere [33]. We chose WhatsApp as our means of data collection because it is the most widely used mobile chat app in Singapore [34] and to protect the health and well-being of participants during the pandemic. Participants were recruited via an online flyer distributed on social media platforms, including Facebook, Twitter, and Instagram. Eligibility criteria were aged 21 years or older and being a Singapore citizen or permanent resident at the point of recruitment. Assuming varying levels of technological savvy and their impacts on group dynamics, participants were purposively recruited by strata according to the following age categories: 21-30 years, 31-40 years, 41-50 years, and 51 years and above. Within each group, we ensured a mix of participants of varying ethnicity, gender, and educational attainment. A summary of participant demographics for the study can be found in Table 1.

Table 1. Summary of participant demographics.

Variables	Participants, n
Age (years)	
21-30	16
31-40	16
41-50	16
≥51	16
Gender identity	
Female	27
Male	20
Another gender	1
Race	
Chinese	42
Malay	2
Indian	2
Another race	2
Formal education attainment	
Secondary and below	1
Preuniversity	13
University	34

Design of FGDs

We conducted two FGDs per age strata, totaling eight FGDs with 6-8 participants per group and 60 participants overall. Each FGD was led by a main facilitator with two observers present. FGDs were conducted over 5 consecutive days, with new discussion topics introduced daily. Topics covered included knowledge and perceptions of COVID-19 and attitudes toward varying information sources.

We conducted both synchronous (ie, all participants were required to be online at a specific time for a specific duration) and asynchronous (ie, participants could reply at their convenience over the course of the day) sessions. Synchronous discussions of approximately 2 hours each were held on the first and last days of the study; asynchronous discussions were held on the second, third, and fourth days. For synchronous sessions, we prepared a list of questions that we had asked participants consecutively during the stipulated timing of 2 hours and participants were expected to stay online to participate at that time. For days that involved asynchronous participation, participants were told that questions would be posed from 9 AM all the way up until 6 PM that day, and that they could choose to answer at any point.

All chat transcripts, including media files such as photos, videos, and memes, were directly downloaded from the researchers' WhatsApp mobile apps and stored in a secure, password-protected location accessible only to research team members. Participants were reimbursed SG $50 for their time (approximately US $35).

Ethics Considerations

Ethics approval was obtained from the Saw Swee Hock School of Public Health Department Ethics Review Committee (SSHSPH-014). All participants provided documented informed consent before participating and completed a demographic questionnaire to indicate their interest to participate.

Data Analysis

Data were analyzed by the lead author, adopting Braun and Clarke's [35,36] six steps of reflexive thematic analysis. Both semantic and latent codes and themes were derived from the data without a pre-existing framework. Following the first two stages of familiarization and coding procedures in classic thematic analysis, the lead author noted clear patterns in how participants discussed and privileged varying forms of health information in the COVID-19 context. The lead author discussed these themes with coauthors to ensure that codes and developing themes were adequately fleshed out and authentic to the data generated [37]. Next, the lead author continued to code according to thematic analysis procedures, drawing links between various constructions and interpretations of health information and the factors underpinning them. At this stage, a typology of interpretation was developed by grouping types along two broad axes based on the perceived trustworthiness of an information source and perceived veracity of health information. All data were organized and analyzed using NVivo 11 software (QSR International Pty Ltd).

Data Quality and Trustworthiness

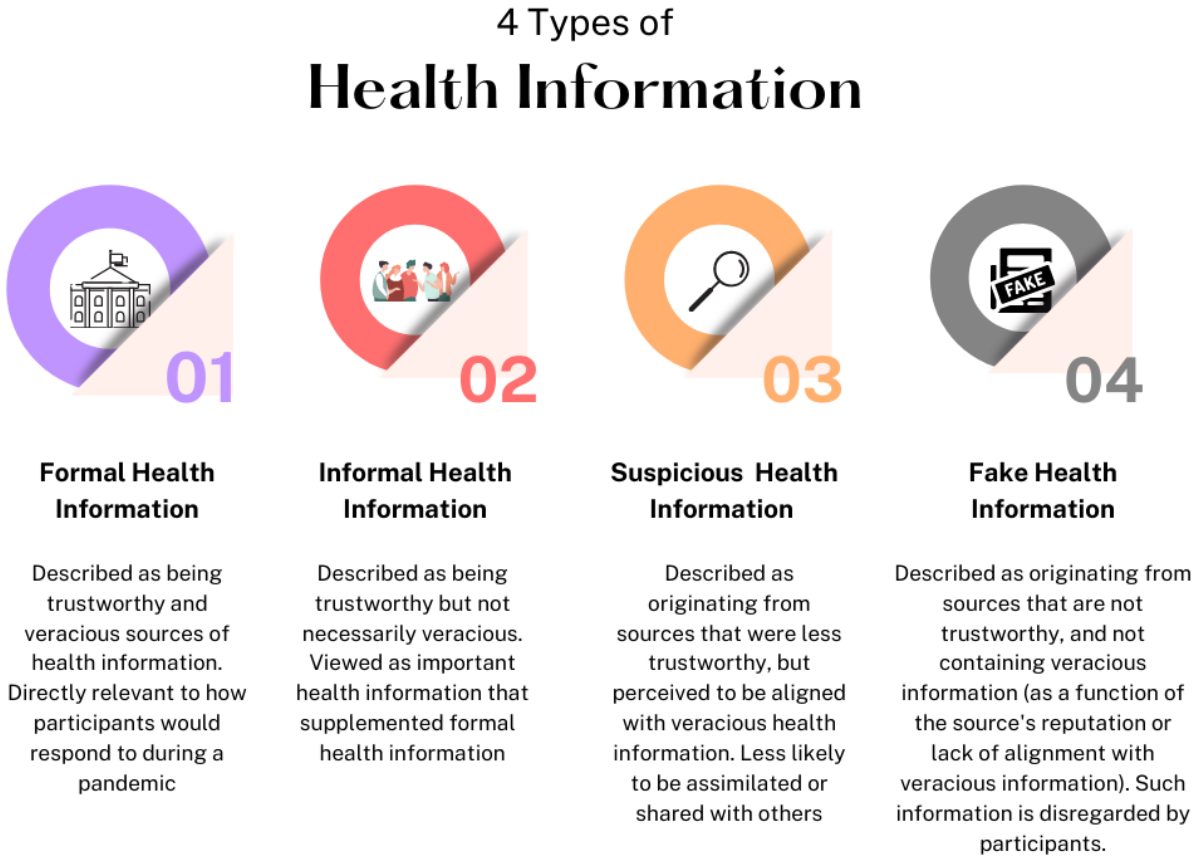
We took several steps to improve the quality and trustworthiness of the data generated in this study. We worked to minimize biases around social desirability or discomfort with sharing by reiterating discussion ground rules on confidentiality, safety, respect, and voluntary participation on a daily basis. To reassure participants of our authenticity, we ensured that our WhatsApp profile pictures used a standardized template featuring a clear photo of our faces and our institutional affiliations. We also created separate "field notes" WhatsApp chat groups for the research team, which we used to reflect on group dynamics and themes in real time, which helped inform probes and prompts for more information and deepen our understanding of participants' interpretations of health information.

Results

Overview

Based on our analysis of responses in our FGDs, we generated four types of health information, including *formal health information*, *informal health information*, *suspicious information*, and *fake health information*. These are summarized in Figure 1.

Figure 1. Summary of four types of health information interpreted by participants.



Formal Health Information

Participants highlighted how they determined what was formal COVID-19 information through an understanding that such information was both trustworthy and veracious. Trustworthiness was premised largely on familiarity and past knowledge of an information source. Veracity was described by participants as being closely aligned with facts from "official" information sources. Participants described that formal health information around COVID-19 was directly relevant to how they would respond to the ongoing pandemic.

Participants described how they depended on trustworthy information sources such as local news channels and the national newspaper, The Straits Times, for formal COVID-19 information. Participants mentioned that the information source was considered trustworthy based on an understanding of its situation within a national regulatory information framework. This trustworthiness of an information source for formal COVID-19 information was reflected in participants' beliefs that while some forms of reporting by the same news outlet may lack transparency due to censorship guidelines in Singapore, they could depend on said information source to report based on public interest. Participants also perceived such information to be accurate based on their understanding that it would be fact-checked. One participant described this as follows:

> *ST [The Straits Times] is a local news and it is under the supervision of IDA [Information Development Authority of Singapore], I know journalists check their facts and they are reviewed by an editor before it is published. […] The only thing is that there might be some info that is not shared for some reason (eg, not to alarm the public)*

Other participants displayed slightly more critical attitudes toward the same information source but were able to distinguish between its position as a trustworthy information source for COVID-19 health information and its political intent. Similarly, they viewed the information source to be "factual and accurate" despite their preconceived notions of its role as state-controlled media:

> *Because it's ST [The Straits Times], I trust that the article is factual and accurate, but sometimes I may be skeptical about the intent of ST articles. I guess there's this general perception that ST is state-controlled media, so their articles may be biased towards the government or certain positions or promoting messages that are favorable to the government. But not that there's anything very wrong about that. I will still read ST articles regardless. […] Yes, I read ST articles quite regularly, especially for local news like updates on the COVID-19 situation and related COVID-19 measures in Singapore.*

Additionally, participants described how in the absence of such an understanding, they could fact-check on their own accord and rely on such information's congruence with other information sources to determine veracity. Participants illustrated this in response to the facilitator's prompt for participants' sources of knowledge on the symptoms of COVID-19:

> *I first heard about it through credible news articles - but I've also googled to see what MOH [Ministry of Health] had to say. and I've checked the government resource (Singapore COVID-19 Symptom Checker [38])* [Participant 1]

> *From news media, but lately, just to prevent my drowning from all the news sites, I just follow CNA [Channel News Asia] and the government's information.* [Participant 2]

Informal Health Information

Participants described informal health information as originating from trustworthy sources, but may contain information that was not perceived to be veracious. Participants discussed how trustworthy information sources, including health information from traditional Chinese medicine practices, religious healing, or other informal sources of health information, served as important information that supplemented formal sources of health information. Participants reported relying on such forms of information as supplementary means of protecting oneself, rather than treating it as formal COVID-19 information.

Participants in the FGDs for older participants offered insight into how some information sources of health information were perceived to be trustworthy even if they did not provide veracious COVID-19 health information. When provided an image prompt depicting the use of onions, turmeric, and ginger to ward off COVID-19, participants discussed how health information from sources such as traditional Chinese medicine could be viewed as trustworthy, even though it was not perceived as being aligned with veracious information on COVID-19:

> *Yes, I heard onion is good even before this COVID-19. This has been circulating for a while already. I don't believe it works this way but I can understand why it is being circulated. They are known to have anti-inflammatory properties when eaten so people might skew this info.* [Participant 1]

> *I believe they [Chinese people] know better than me. This is not supported by scientific research. […] Being elderly, more familiar with traditional Chinese medicine (TCM), but especially Chinese.* [Participant 2]

Other participants cited how friends served as trustworthy sources of information, even though they could not ascertain the veracity of such health information in the context of COVID-19 prevention. When prompted with the same picture from above, a participant from a separate FGD highlighted the following: "I heard onions do kill germs, and friends do practice it. But how effective, I don't know."

Other participants also discussed the use of a "silver ion spray"; one participant shared information about and promoted the product, citing its antiviral properties, even though such information was not specific to COVID-19:

> *Never seen or heard. But heard of silver ion spray. Please watch this video with a high-resolution camera showing how the virus can be spread. Don't bother about the Japanese language. The images are enough to understand. [Video attachment of silver ion spray product]. It's an alcohol-free antivirus spray that can kill up to 99.9% [of viruses]. It's safe to be sprayed into the eye as well.*

Suspicious Health Information

Participants described suspicious health information as typically originating from less trustworthy information sources, even though it might be consistent with known, veracious

COVID-19–related information. Such information was typically not viewed as a favorable information source that participants would assimilate or share with others. One participant pointed out how despite factual reporting of COVID-19 information from certain sources, they would "take it with a pinch of salt" due to perceived political biases of such publications:

> *We even discount [South China Morning Post] as credible. We need to analyze whether certain posts are politically motivated. News outlets like Fox News may be official and credible but I will take it with a pinch of salt.*

When asked to discuss and compare three articles by various news outlets on the same topic, participants highlighted that their levels of trust in that same piece of health information varied depending on the source of the news article. One participant highlighted how it was difficult to measure trustworthiness among news articles and preferred to stick to government bodies as an authority for formal COVID-19 information. Participants also pointed out that a media outlet's trustworthiness would also depend on the subject matter, and that they would trust varying sources depending on whether they were reporting on certain topics or geographies. A discussion by several participants highlights this nuance:

> *I'll trust ST [The Straits Times] more because I've read articles from BBC [British Broadcasting Corporation] and SCMP [South China Morning Post], especially that latter, that have a very biased view against Singapore. Since ST is more towards the local context, as a Singaporean I'll trust it more.* [Participant 1]
>
> *Depending on the locations that are being written about, I would trust different sources. For example, if I wish to find out about local updates, I would be inclined to read Straits Times instead.* [Participant 2]

Fake Health Information

Participants described how fake health information usually came from untrustworthy sources and could be viewed as not being veracious through two mechanisms. First, it could be determined as a function of trustworthiness, in that information was not trustworthy owing to its historical and consistent lack of fact-checking. Therefore, information would be viewed as inaccurate regardless of its content. Second, information from such sources did not align or comport with veracious COVID-19 information propagated by official sources. In general, participants disregarded such forms of health information.

Broadly, participants across most FGDs described how forwarded WhatsApp messages from relatives were regarded as untrustworthy, and therefore contained unverified information on COVID-19. Participants from one FGD forwarded a text message to the group from a relative on the use of food to protect oneself from COVID-19, and subsequently commented on its trustworthiness:

> *I didn't take it seriously because there was no trustworthy source cited.* [Participant 1]
>
> *Recently I couldn't find vinegar in the supermarket. I asked my mummy's circle [a motherhood-centered social group] why vinegar is OOS [out of stock]. According to them, some elderly believe drinking vinegar can protect themselves from COVID-19. Btw [by the way], I just need it for my gyoza [Japanese dumpling].* [Participant 2]

By contrast, participants in a separate FGD focused on COVID-19 prevention measures in their discussion on determining fake health information. One participant had forwarded a video that was circulating on WhatsApp of a police officer chatting with a civilian on the sidewalk next to a police van and issuing a fine. Another participant responded that it was untrustworthy by virtue of how it was shared (ie, through WhatsApp). However, while participants viewed forwarded messages as an untrustworthy source, they also described how they would discern if such information aligned with their knowledge of accurate and formal COVID-19 information. Specifically, the original poster approached this as *suspicious health information* instead and called this a "half-truth" due to the lack of trustworthiness of the source, although aligning with their understanding around current COVID-19 movement control measures:

> *[This video] Definitely, sent through WA [WhatsApp] groups.* [Participant 1]
>
> *I saw that too. I feel like it's some sort of a half-truth, around people being issued notices [for flouting safe distancing measures]. I don't know if the actual fines are legit [legitimate].* [Participant 2]
>
> *I think the fines are for people who do not observe safe distancing. The authorities encourage people to wear masks now even if well but it is not mandatory. So yeah, such forwarded message creates confusion.* [Participant 1]

Discussion

Principal Findings

Our study found that participants interpreted four types of health information, namely *formal health information*, *informal health information*, *suspicious health information*, and *fake health information*. These forms of health information were determined as a function of the perceived trustworthiness of a given information source and the perceived veracity of information. We discuss the implications of these findings in the context of existing studies on health literacy and approaches to dealing with the COVID-19 infodemic.

We found that participants were able to distinguish between formal and informal sources of health information. Determining if health information belonged to these respective categories would involve receiving information from trustworthy sources of information and assessing if it is veracious or not, which may be a given for certain sources of formal information or requires further fact-checking. For study participants, supplementary health information sources played an important role in providing additional perceived protections from COVID-19, although participants recognized that such information was not official

or they were unaware of the mechanisms through which it would positively impact COVID-19 prevention efforts. This was especially salient for the older age groups, who discussed how certain forms of trustworthy sources, including alternative healing, traditional Chinese medicine, or frameworks of cultural beliefs, provided resources to supplement their COVID-19 prevention behaviors. Studies in other settings have found that complementary, traditional, or alternative forms of medicines have played a strong role in the alleged treatment and prevention of COVID-19 [39-41], despite the lack of rigorous randomized controlled trials to underpin such evidence [42,43].

The trustworthiness of a source played a key role in determining if participants would assimilate or share certain forms of information. In general, both suspicious and fake health information led to participants having reservations around assimilating and sharing such information, even if they knew that the information had some "truth" in it due to existing knowledge of formal and accurate COVID-19 information. Past studies show that trust in a particular source plays a strong role in eventual health information–seeking behaviors [44,45]. In turn, those with higher levels of health literacy tended to trust information from health care professionals and were less likely to do so for information from social media, celebrities, and friends [46]. This comports with our finding that older adults, who have been found in multiple settings to have limited health literacy [47-49], may also rely on complementary or alternative sources of health information to inform their COVID-19 prevention efforts. Several scholars have also reviewed the antecedents of trust for online health information [50,51], while others have proposed sociodemographic correlates of trust for varying sources of offline health information [52,53].

Overall, our findings indicate that individuals may label the same piece of health information differently depending on their perceptions of a source's trustworthiness and veracity of health information. This finding aligns with research that points out a shift toward a "posttruth" era, one that situates truthful information as at risk of being undermined by "alternative facts" or misinformation [54]. The findings of this study are also consistent with past scholarship that posits how the dynamics of belief formation and the definition of truth are being contested in contemporary societies [55], which may impact health and media literacy in society.

Strengths and Limitations

We identified three key study strengths. First, we generated a dynamic framework to understand how people approach or interpret health information in the context of an infodemic. The typology provides opportunities for policymakers to propose interventions and communication strategies that may enable effective and rapid communication in a pandemic, while acknowledging the varied ways in which individuals interpret information. Second, WhatsApp proved to be an appropriate platform for discussions on health literacy in times of an infodemic, as participants were able to share and forward information and media that helped initiate dynamic discussions on their approaches to determining the trustworthiness and/or veracity of such information. Third, WhatsApp-based FGDs allowed us to continue with the study despite the ongoing movement control restrictions that were progressively implemented midway through the study, thus allowing us to retain the integrity of our proposed methods across all groups while keeping our participants safe.

We are also mindful of several limitations. First, the use of WhatsApp FGDs did not allow facilitators to make sense of nonverbal cues such as the use of body language, which might have diminished rapport between participants and researchers. Furthermore, the lack of tone that would normally be present in verbal communication meant that some meanings could have been misconstrued by the facilitators. Nevertheless, our team members were able to analyze emojis, images, or stickers to gain more context in lieu of such cues. These provided additional context for tone when analyzing participants' responses:

> *This is the most prolonged WhatsApp conversation I've been a part of*  *[…]* [Participant 1]
>
> *Thank you for the great insights! Have a great week ahead*  *[…]* [Participant 2]
>
> *Not that I remember*  [Participant 3]

Furthermore, having experience within the team in employing this method, steps were taken to mitigate such methodological shortcomings, as articulated in the Methods section [33].

Conclusion

We conclude with two specific recommendations for health authorities and policymakers to enhance effective communications during a health infodemic such as during the COVID-19 pandemic. These recommendations aim to intervene on levels of perceived trust in health information sources and the perceived veracity of health information. In the context of the trustworthiness of health information sources, we recommend policies that shape norms and build trust in select sources of information to combat a glut of health information. The nature of trust and trustworthiness of sources is complex, and draws on participants' past experiences and literacy around a particular source's attributes and authority. Scholarly work on the antecedents of trust for online health information might serve as useful starting points for interventions aimed at promoting trust in key sources of information. Such efforts should be implemented well before times of crises and infodemics.

Second, we recommend acknowledging nuances between formal, evidence-based information and information on alternative or complementary medicine or treatments, and implementing an information framework that distinguishes between, yet supports both. Our results suggest that the recognition of supplementary health information may not necessarily be harmful, given that such information may be used to complement formal health information in times of a pandemic. However, this presupposes a strong understanding of what comprises formal and veracious health information and the information architecture that supports such interpretations. Upon establishing key sources of veracious information, individuals may be better equipped to distinguish what *must* be done to protect themselves in a pandemic, relative to what *can*

be done as a means of supplementing formal COVID-19 prevention measures.

Acknowledgments

The publication of this study was generously funded by the Saw Swee Hock School of Public Health, National University of Singapore and the National University Health System. We thank all participants for their contributions to, and our colleagues for their inputs and support for the study.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2022-06-16]
2. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
3. Weekly epidemiological update on COVID-19 - 20 April 2022. World Health Organization. 2022. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---20-april-2022 [accessed 2022-04-26]
4. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
5. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
6. Munich Security Conference. World Health Organization. 2020. URL: https://www.who.int/director-general/speeches/detail/munich-security-conference [accessed 2021-02-22]
7. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health 2020 May;5(5):e249-e250 [FREE Full text] [doi: 10.1016/S2468-2667(20)30086-4] [Medline: 32302535]
8. Spring H. Health literacy and COVID-19. Health Info Libr J 2020 Sep 16;37(3):171-172 [FREE Full text] [doi: 10.1111/hir.12322] [Medline: 32672399]
9. Abel T, McQueen D. Critical health literacy and the COVID-19 crisis. Health Promot Int 2020 Dec 01;35(6):1612-1613 [FREE Full text] [doi: 10.1093/heapro/daaa040] [Medline: 32239213]
10. Okan O, Bollweg TM, Berens E, Hurrelmann K, Bauer U, Schaeffer D. Coronavirus-related health literacy: a cross-sectional study in adults during the COVID-19 infodemic in Germany. Int J Environ Res Public Health 2020 Jul 30;17(15):5503 [FREE Full text] [doi: 10.3390/ijerph17155503] [Medline: 32751484]
11. Earnshaw V, Eaton LA, Kalichman SC, Brousseau NM, Hill EC, Fox AB. COVID-19 conspiracy beliefs, health behaviors, and policy support. Transl Behav Med 2020 Oct 08;10(4):850-856 [FREE Full text] [doi: 10.1093/tbm/ibaa090] [Medline: 32910819]
12. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J 2021 Jun 12;38(2):143-149 [FREE Full text] [doi: 10.1111/hir.12320] [Medline: 32657000]
13. Moscadelli A, Albora G, Biamonte MA, Giorgetti D, Innocenzio M, Paoli S, et al. Fake news and Covid-19 in Italy: results of a quantitative observational study. Int J Environ Res Public Health 2020 Aug 12;17(16):5850 [FREE Full text] [doi: 10.3390/ijerph17165850] [Medline: 32806772]
14. Dash S, Parray AA, De Freitas L, Mithu MIH, Rahman MM, Ramasamy A, et al. Combating the COVID-19 infodemic: a three-level approach for low and middle-income countries. BMJ Glob Health 2021 Jan 29;6(1):e004671 [FREE Full text] [doi: 10.1136/bmjgh-2020-004671] [Medline: 33514596]
15. Seng JJB, Yeam CT, Huang CW, Tan NC, Low LL. Pandemic related health literacy – a systematic review of literature in COVID-19, SARS and MERS pandemics. medRxiv. 2020 May 11. URL: https://www.medrxiv.org/content/10.1101/2020.05.07.20094227v1 [accessed 2022-11-07]
16. Riiser K, Helseth S, Haraldstad K, Torbjørnsen A, Richardsen KR. Adolescents' health literacy, health protective measures, and health-related quality of life during the Covid-19 pandemic. PLoS One 2020 Aug 28;15(8):e0238161 [FREE Full text] [doi: 10.1371/journal.pone.0238161] [Medline: 32857806]
17. Nguyen HT, Do BN, Pham KM, Kim GB, Dam HT, Nguyen TT, et al. Fear of COVID-19 Scale-Associations of its scores with health literacy and health-related behaviors among medical students. Int J Environ Res Public Health 2020 Jun 11;17(11):4164 [FREE Full text] [doi: 10.3390/ijerph17114164] [Medline: 32545240]
18. Ashrafi-Rizi H, Kazempour Z. Information typology in coronavirus (COVID-19) crisis; a commentary. Arch Acad Emerg Med 2020;8(1):e19 [FREE Full text] [Medline: 32185370]

19. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
20. Ferrara E. What types of COVID-19 conspiracies are populated by Twitter bots? First Monday 2020 May 19;25(6):10633 [FREE Full text] [doi: 10.5210/fm.v25i6.10633]
21. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
22. Oates C. The use of focus groups in social science research. In: Burton D, editor. Research training for social scientists. Thousand Oaks, CA: Sage Publishing; 2000:186-195.
23. Stewart K, Williams M. Researching online populations: the use of online focus groups for social research. Qual Res 2016 Aug 17;5(4):395-416. [doi: 10.1177/1468794105056916]
24. Tuttas CA. Lessons learned using Web conference technology for online focus group interviews. Qual Health Res 2015 Jan 05;25(1):122-133. [doi: 10.1177/1049732314549602] [Medline: 25192765]
25. Brüggen E, Willems P. A critical comparison of offline focus groups, online focus groups and E-Delphi. Int J Market Res 2018 Apr 25;51(3):1-15. [doi: 10.1177/147078530905100301]
26. Schneider SJ, Kerwin J, Frechtling J, Vivari BA. Characteristics of the discussion in online and face-to-face focus groups. Soc Sci Comput Rev 2016 Aug 18;20(1):31-42. [doi: 10.1177/089443930202000104]
27. Woodyatt CR, Finneran CA, Stephenson R. In-person versus online focus group discussions: a comparative analysis of data quality. Qual Health Res 2016 May 01;26(6):741-749. [doi: 10.1177/1049732316631510] [Medline: 26935719]
28. Lijadi AA, van Schalkwyk GJ. Online Facebook focus group research of hard-to-reach participants. Int J Qual Meth 2015 Dec 11;14(5):160940691562138. [doi: 10.1177/1609406915621383]
29. Abrams KM, Wang Z, Song YJ, Galindo-Gonzalez S. Data richness trade-offs between face-to-face, online audiovisual, and online text-only focus groups. Soc Sci Comput Rev 2014 Jun 03;33(1):80-96. [doi: 10.1177/0894439313519733]
30. Chen J, Neo P. Texting the waters: an assessment of focus groups conducted via the WhatsApp smartphone messaging application. Method Innov 2019 Nov 04;12(3):205979911988427. [doi: 10.1177/2059799119884276]
31. Colom A. Using WhatsApp for focus group discussions: ecological validity, inclusion and deliberation. Qual Res 2021 Jan 21;22(3):452-467. [doi: 10.1177/1468794120986074]
32. Yip W, Ge L, Ho AHY, Heng BH, Tan WS. Building community resilience beyond COVID-19: The Singapore way. Lancet Reg Health West Pac 2021 Feb;7:100091 [FREE Full text] [doi: 10.1016/j.lanwpc.2020.100091] [Medline: 33521745]
33. Neo PH, Lim JM, Tan RK, Ong SE. Using WhatsApp Focus Group Discussions for Qualitative Data Collection: Exploring Knowledge, Attitudes, and Perceptions of COVID-19 in Singapore. Int J Qual Methods 2022 Apr 10;21:16094069221090355 [FREE Full text] [doi: 10.1177/16094069221090355] [Medline: 35437433]
34. Digital 2021 Singapore. We Are Social. 2021. URL: https://wearesocial.com/sg/blog/2021/01/digital-2021-singapore/ [accessed 2022-11-07]
35. Braun V, Clarke V. Successful qualitative research: a practical guide for beginners. London: Sage Publishing; 2013.
36. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
37. Morrow SL. Quality and trustworthiness in qualitative research in counseling psychology. J Counsel Psychol 2005 Apr;52(2):250-260. [doi: 10.1037/0022-0167.52.2.250]
38. Singapore COVID-19 Symptom Checker. URL: https://www.sgcovidcheck.gov.sg/ [accessed 2022-11-07]
39. Li Q, Wang H, Li X, Zheng Y, Wei Y, Zhang P, et al. The role played by traditional Chinese medicine in preventing and treating COVID-19 in China. Front Med 2020 Oct 10;14(5):681-688 [FREE Full text] [doi: 10.1007/s11684-020-0801-x] [Medline: 32651936]
40. Charan J, Bhardwaj P, Dutta S, Kaur R, Bist SK, Detha MD, et al. Use of complementary and alternative medicine (CAM) and home remedies by COVID-19 patients: a telephonic survey. Indian J Clin Biochem 2021 Jan 31;36(1):108-111 [FREE Full text] [doi: 10.1007/s12291-020-00931-4] [Medline: 33162692]
41. Weeks J. Call to Action: announcing the traditional, complementary and integrative health and medicine COVID-19 support registry. J Altern Complement Med 2020 Apr 01;26(4):256-258. [doi: 10.1089/acm.2020.29083.jjw] [Medline: 32250655]
42. Shankar A, Dubey A, Saini D, Prasad CP. Role of complementary and alternative medicine in prevention and treatment of COVID-19: an overhyped hope. Chin J Integr Med 2020 Aug 05;26(8):565-567 [FREE Full text] [doi: 10.1007/s11655-020-2851-y] [Medline: 32761336]
43. Ganguly S, Bakhshi S. Traditional and complementary medicine during COVID-19 pandemic. Phytother Res 2020 Dec 12;34(12):3083-3084 [FREE Full text] [doi: 10.1002/ptr.6828] [Medline: 32785960]
44. Yang Q, Chen Y, Wendorf Muhamad J. Social Support, Trust in Health Information, and Health Information-Seeking Behaviors (HISBs): a study using the 2012 Annenberg National Health Communication Survey (ANHCS). Health Commun 2017 Sep 02;32(9):1142-1150. [doi: 10.1080/10410236.2016.1214220] [Medline: 27589249]
45. Hou J, Shim M. The role of provider-patient communication and trust in online sources in internet use for health-related activities. J Health Commun 2010 Dec 10;15(Sup 3):186-199. [doi: 10.1080/10810730.2010.522691] [Medline: 21154093]

46. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018 Aug 30;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
47. Berens E, Vogt D, Messer M, Hurrelmann K, Schaeffer D. Health literacy among different age groups in Germany: results of a cross-sectional survey. BMC Public Health 2016 Nov 09;16(1):1151 [FREE Full text] [doi: 10.1186/s12889-016-3810-6] [Medline: 27829401]
48. von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of British adults. J Epidemiol Community Health 2007 Dec 01;61(12):1086-1090 [FREE Full text] [doi: 10.1136/jech.2006.053967] [Medline: 18000132]
49. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec 05;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
50. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
51. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun 16;22(2):355-369 [FREE Full text] [doi: 10.1177/1460458214559432] [Medline: 25518944]
52. Peterson E, Chou WYS, Kelley DE, Hesse B. Trust in national health information sources in the United States: comparing predictors and levels of trust across three health domains. Transl Behav Med 2020 Oct 08;10(4):978-988 [FREE Full text] [doi: 10.1093/tbm/ibz066] [Medline: 31116400]
53. Jackson DN, Peterson EB, Blake KD, Coa K, Chou WS. Americans' trust in health information sources: trends and sociodemographic predictors. Am J Health Promot 2019 Nov 23;33(8):1187-1193. [doi: 10.1177/0890117119861280] [Medline: 31337226]
54. McIntyre L. Post-Truth. Cambridge, MA: MIT Press; 2018.
55. Waisbord S. Truth is what happens to news. Journal Stud 2018 Jul 06;19(13):1866-1878. [doi: 10.1080/1461670x.2018.1492881]

Abbreviations

DORSCON: Disease Outbreak Response System Condition
FGD: focus group discussion
WHO: World Health Organization

Edited by A Kushniruk; submitted 05.05.22; peer-reviewed by M Popovic, SF Tsao; comments to author 17.08.22; revised version received 05.09.22; accepted 10.09.22; published 22.11.22.

Please cite as:
Tan RKJ, Lim JM, Neo PHM, Ong SE
Reinterpretation of Health Information in the Context of an Emerging Infectious Disease: A Digital Focus Group Study
JMIR Hum Factors 2022;9(4):e39312
URL: https://humanfactors.jmir.org/2022/4/e39312
doi:10.2196/39312
PMID:36099011

Original Paper

International Collaboration to Develop a Remote Monitoring Web App for COVID-19 Patients in Armenia: Design and Development With Agile Methodology

Abu Sikder[1*], BA; James Dickhoner[1*], MD; Lynn Kysh[1*], MLS; Lusine Musheghyan[2*], MSW, MPH; Shant Shekherdimian[3*], MD; Barry Levine[4*], PhD; Juan Espinoza[1,5*], MD

[1]Innovation Studio, Children's Hospital Los Angeles, Los Angeles, CA, United States

[2]Turpanjian College of Health Sciences, American University of Armenia, Yerevan, Armenia

[3]Division of Pediatric Surgery, University of California, Los Angeles, Los Angeles, CA, United States

[4]Department of Computer Science, San Francisco State University, San Francisco, CA, United States

[5]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Juan Espinoza, MD
Department of Pediatrics
Keck School of Medicine
University of Southern California
4650 Sunset Blvd
Los Angeles, CA, 90027
United States
Phone: 1 323 660 2450
Email: jespinoza@chla.usc.edu

Abstract

Background: COVID-19 has led to over 500 million cases and 6.2 million deaths around the world. Low- and middle-income countries (LMICs) like Armenia face unique infrastructure, financial, and capacity challenges that in many cases result in worse outcomes. Health care facilities across Armenia experienced a shortage of resources, including hospital beds and oxygen, which was further exacerbated by the war with neighboring Azerbaijan. Without a framework for home-based care, health care facilities were severely strained by COVID-19 patients who had prolonged oxygen requirements but were otherwise clinically stable.

Objective: This paper describes our approach to establishing an international collaboration to develop a web app to support home monitoring of patients with COVID-19 with persistent oxygen requirements.

Methods: The app was developed using a rapid, coordinated, and collaborative approach involving an international group of clinicians, developers, and collaborators. Health screening, monitoring, and discharge forms were developed into a lightweight OpenMRS web app and customized for the local Armenian context.

Results: The software was designed and developed over 2 months using human-centered design and agile sprints. Once live, 5087 patient records were created for 439 unique patients.

Conclusions: This project suggests a promising framework for designing and implementing remote monitoring programs in LMICs, despite pandemic and geopolitical challenges.

(JMIR Hum Factors 2022;9(4):e40110) doi:10.2196/40110

KEYWORDS

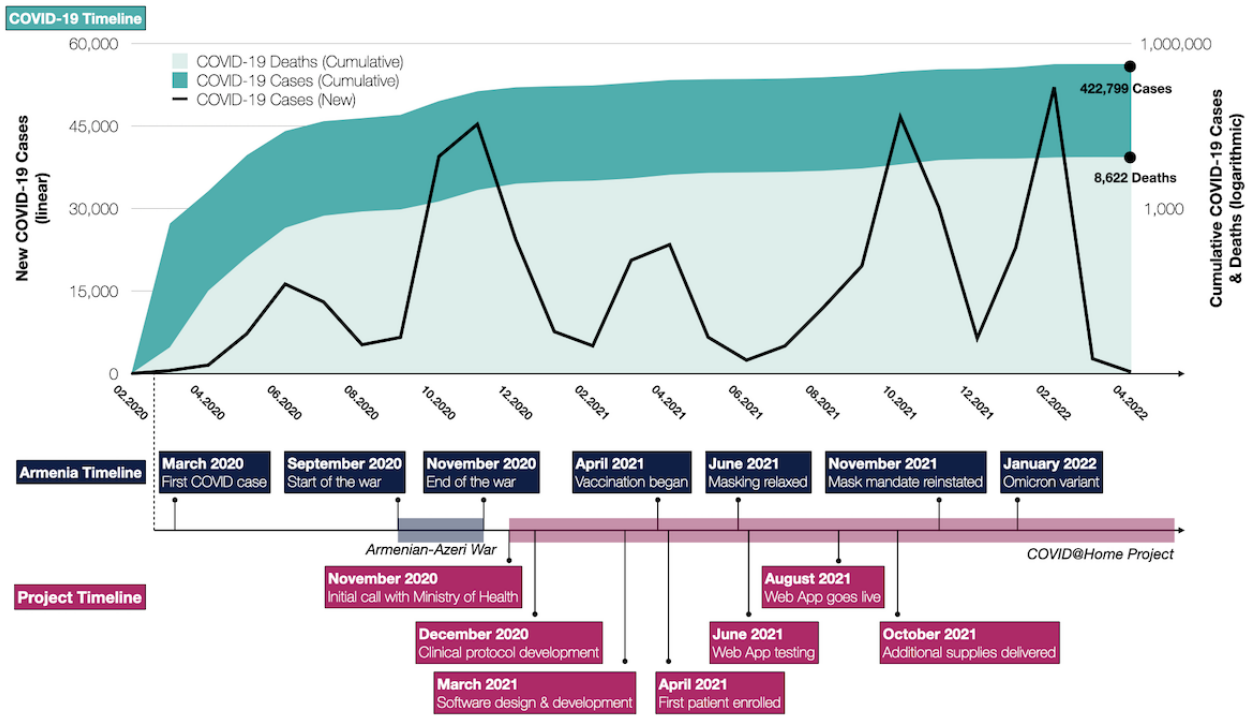
COVID-19; global health; software; mHealth; Armenia; web app; home monitoring; software development; human-centered design; remote monitoring; patient care

Introduction

The rapid transmissibility and pathogenicity of COVID-19 has led to over 500 million cases and 6.2 million deaths worldwide [1-4]. Health care facilities around the world have become overwhelmed through depletion of critical resources, such as hospital beds, mechanical ventilators, and key medications [5]. Low- and middle-income countries (LMICs) face unique infrastructure, financial, and capacity challenges that in many cases result in worse outcomes [6].

Armenia, an LMIC situated in western Asia with a population of 2.98 million, had among the highest prevalence of COVID-19 for several months [1,7,8]. As a post-Soviet nation operating under the Semashko model's influence, Armenia's health care system is affected by significant fragmentation and specialization of hospitals [8]. The first COVID-19 case in Armenia was identified in March 2020, with numbers increasing substantially in the following months (Figure 1) [8]. There was a shortage of resources, including hospital beds and oxygen, that was further exacerbated by the war with Azerbaijan for the disputed Nagorno-Karabakh region from September to November 2020 [8,9]. The Armenian health care system faced immense pressure during this time, as health care facilities struggled to manage COVID-19 in addition to war casualties and other public health challenges [8]. Without a framework for home-based care, hospitals became crowded with patients with ongoing oxygen requirements who were otherwise clinically stable. This became a significant burden during the second wave of the pandemic in early 2021.

Figure 1. Composite timeline of events in Armenia related to the COVID@Home Project (data obtained from Our World in Data [10], which is licensed under Creative Commons Attribution 4.0 International License [11]).



In response, Armenia's Ministry of Health, the Turpanjian College of Health Sciences of the American University of Armenia, the University of California, Los Angeles (UCLA), and Children's Hospital Los Angeles (CHLA) collaborated to develop a remote monitoring program to enable patients with prolonged oxygen requirements to be safely discharged home. Teams of clinicians, software developers, researchers, and administrative staff from both US- and Armenia-based institutions designed and implemented a home-based remote monitoring program (COVID@home) for low-risk patients tailored to the local Armenian context. This multinational, multidisciplinary collaboration was made possible by long-standing partnerships between these institutions focused on improving child health outcomes, preventive health screenings, supporting postgraduate medical education, implementing health technology innovation, and building local capacity [12-17]. Both UCLA and CHLA have existing institutional and grant-funded priorities to partner with Armenian institutions to improve health and health care delivery.

Remote patient monitoring programs have shown promise in managing low-risk patients while offloading the burden and use of critical resources in health care facilities. Although studies have adapted these programs to manage COVID-19 patients, the literature is limited on their utility and design, and none have been developed for Armenia. Due to cost, health care infrastructure requirements, and underdeveloped regulatory frameworks, there are many challenges with implementing and maintaining these systems in LMICs [15]. A promising and popular solution is OpenMRS, an open-source electronic health record (EHR) platform maintained by a global community of developers [18,19]. For this project, we customized an instance of OpenMRS and developed the COVID@Home web app to support our remote monitoring program in Armenia. Many software development methodologies exist, such as waterfall, spiral, V-shaped, and agile [20]. Due to the rapidly evolving

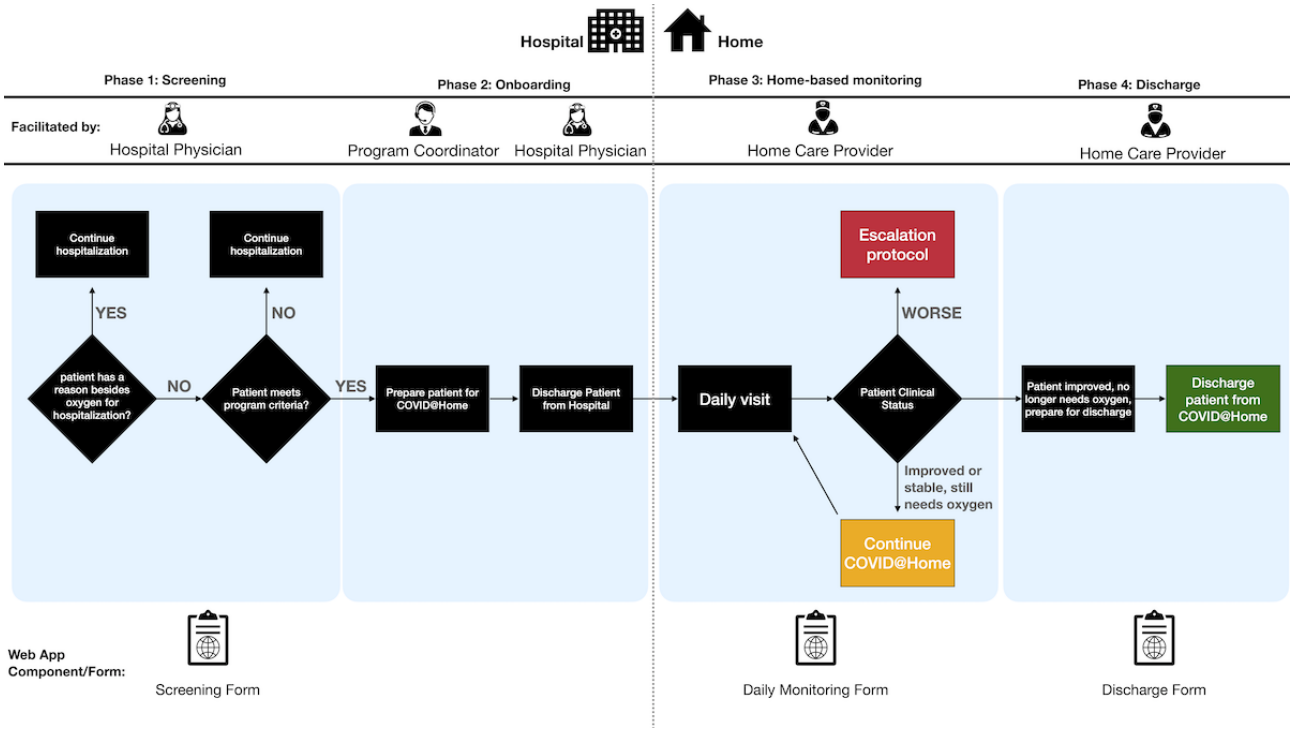
nature of the COVID-19 pandemic, we leveraged an iterative, agile software development methodology in order to be responsive to the realities on the ground [20,21]. This paper describes our specific approach to designing and developing the web app using a collaborative, international, multidisciplinary team amidst unique pandemic and geopolitical challenges. A separate report covering the clinical protocol and outcomes is in preparation.

Methods

Program Overview

The COVID@Home program was designed to provide home monitoring of hospitalized patients with COVID-19 who had an ongoing oxygen requirement but were otherwise clinically stable. The two main components of the program were (1) inpatient screening and (2) at-home management, both of which were facilitated by the web app (Figure 2). Patient screening for program eligibility, which included a home safety evaluation, used the web app and was performed by a project nurse. If eligible, the patient was enrolled and consent was requested. Hospital discharge notes were made using the web app, and the patient was discharged with oxygen supplies and a pulse oximeter, along with appropriate education and instructions. Each day a home care provider (HCP) called the patient to check in on them and collect monitoring information using the web app. Once patients met specific clinical criteria, they were discharged from the program, the oxygen supplies were picked up from their home, and a discharge note was completed using the web app. Patients were transferred to the hospital if their condition worsened.

Figure 2. COVID@Home program overview and workflow with corresponding web app forms.



Patient Population

The target patient population was hospitalized adult COVID-19 patients in Armenia. Clinically, patients were required to have stable vital signs (including blood pressure, heart rate, temperature, and breathing rate), stable glucose levels, and oxygen saturation >93% on no more than 10 L/min. Other eligibility criteria included access to a phone, a safe living environment, and access to electricity. The program was hosted in Yerevan with patients from 7 participating hospitals: Mikayelyan Institute of Surgery, Surb Grigor Lusavorich Medical Center, the National Center for Infectious Diseases, Avan Clinic of Surb Grigor Lusavorich Medical Center, Surb Astvatsamayr Medical Center, Scientific Center of Traumatology and Orthopaedy, and Erebuni Medical Center. Patients lived either in Yerevan or in close proximity to it (<40 km).

Software Development Team

Development of the software involved 3 clinicians, 2 program administrators, 3 software developers, and 1 project manager. All team members resided in either Armenia or the United States.

Ethics Approval

This study received internal review board approval (#CHLA-22-00028) from the CHLA.

Software Development

Human-Centered Design

Human-centered design (HCD) is an empathetic problem-solving technique centered around the experiences and needs of people [22]. It is often used as an iterative and collaborative process that intimately involves stakeholders and end users [22]. Our process began with identifying and

interviewing critical stakeholders and users of the tool, ranging from administrators and leaders to frontline clinical staff. By engaging these individuals early on, we were able to develop trust and gather their feedback at each stage of the development process. Our users identified specific problems and their needs for the proposed tool. We completed several rounds of ideation using low-fidelity mockups for rapid prototyping and gathered feedback from our users to inform the development of our tool. Once the research team and stakeholders arrived at a meaningful design, a functional version was released. We conducted additional interviews with the program managers to ensure the tool was well integrated into our end users' workflows. The app was then localized by professional Armenian translators for both proper language use and cultural competency. Finally, after deploying the tool, we completed usability testing with our users to further refine the tool and ensure it was meeting their needs.

Project Management

The web app was developed using an iterative, agile methodology based on weekly sprint sessions with stakeholders. Each week the core team (program administrators, researchers, and software developers) met over Zoom (Zoom Video Communications) to review and set tasks involving the creation of new features, making updates, fixing bugs, and testing the software. Action items were stored as tickets in JIRA, a software management program (Atlassian). Updates were recommended by core members and end users after testing and reviewing the features or code. Stakeholders were actively consulted after milestones to obtain feedback and improve on the design and functionality of the software as needed.

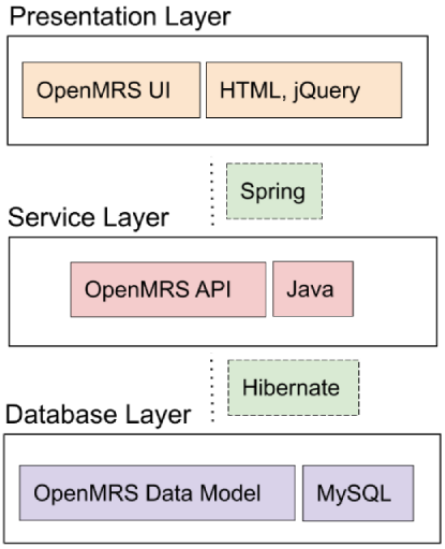
Tools and Technology

OpenMRS is a web-based application built using Spring MVC (a web app framework) that uses a custom application programming interface (API) to interact with the data model [23]. It relies on a concept dictionary based on the Columbia International eHealth Laboratory for storing and retrieving patient data [24]. Another important design feature of OpenMRS is the use of modules that have full access to the OpenMRS system, including the database, API, and front end. The modules are used to expand or customize features of the core application.

Architecture

The web app architecture includes a presentation layer (PL) with HTML, jQuery, and other front-end technologies (eg, React), a service layer (SL) with Java, and a database layer (DL) with MySQL (Figure 3). Spring MVC helps interface the PL and SL by following the model-view-controller design pattern [25,26]. Hibernate is used as an object relational mapper [27]. All the various layers can be accessed and modified using modules via direct access to the API. The web app also uses custom-built OpenMRS add-on modules that interact with all 3 layers of the core system. The modules provide additional web pages with associated SL code and database tables.

Figure 3. Architecture of the COVID@home OpenMRS platform. API: application programming interface; UI: user interface.



Data Collected

Forms

Three forms were developed by the clinical team and integrated into the web app: (1) a screening form completed by hospital and program staff, (2) a daily monitoring form completed by the HCPs, and (3) a discharge form completed by the HCPs once a patient was discharged from the program. Each form contained a distinct set of questions about the participant's health, demographics, vital signs, and clinical history. Both English and Armenian versions of the forms were made available.

Workflow

The HCPs updated the forms directly in the web app for screening, health monitoring, and discharge. In addition, hard copies were made available, and the information was re-entered manually in the web app system at a later time. The data were then made available via reports created within the OpenMRS system.

Results

HCD Process

The HCD process lasted 2 months, with design meetings involving US- and Armenian-based physicians, engineers, and project managers that were hosted over Zoom. Deliverables included a complete wireframe, a spreadsheet with Armenian language translations, and a document with user acceptance testing (UAT) feedback. Deliverables were reviewed by Armenian clinicians for suitability for use in Armenia. Comments from the UAT helped improve the user interface and user experience, such as by adding clicks, improving navigation, and adding a function to validate form input. Hard-copy backups were included at the suggestion of users.

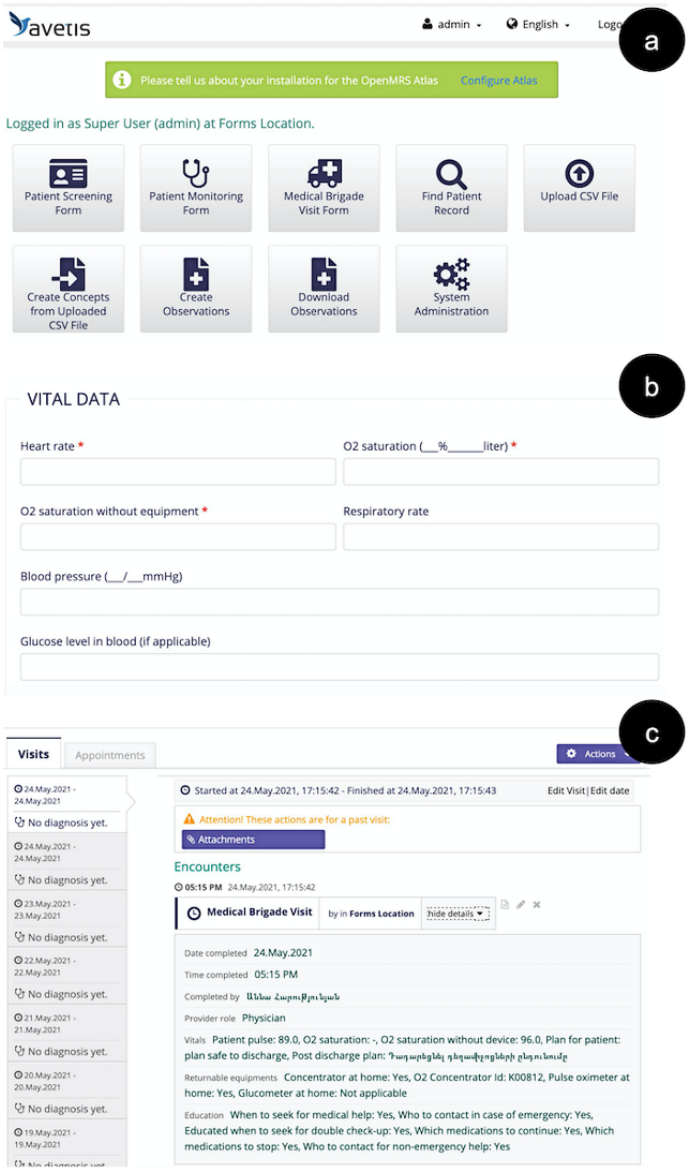
Development Logistics

Six biweekly meetings were held with the software developers to finalize the platform. Meeting times were adjusted to better accommodate the fluctuating and busy schedules of the clinical staff. Using the initial deliverables from HCD, the software developers created a demonstration version of the app that was validated by stakeholders through trial and error. Over 200 JIRA tickets were completed during development and testing. The final version was rolled out to end users and implemented by local stakeholders with ancillary services from US colleagues.

Software Components: Features and Functionality

Five customized web pages were created: a login page, a home page, and 3 screening forms (Figure 4). Four OpenMRS modules were developed for inserting comma-separated value (CSV) data from hard-copy forms, exporting form data, editing participant status, and finding records. There were more than 100 new OpenMRS concepts created to accommodate each data point in both Armenian and English. Concepts were integrated using 2 Groovy scripts: CSV to concepts and CSV to observations. Custom drug classes were also included using 2 Groovy scripts: CSV to drug classes and concept from drug classes. A total of 52% of the code was written in Java and 47% in Groovy.

Figure 4. The OpenMRS front end for (a) COVID@home, with forms, modules, and admin options; (b) the form data entry page; and (c) an example of a completed patient record.



Platform Implementation and Utilization

The COVID@Home program started enrolling patients in April 2021 (Figure 1). During the first few months, while the web app was still being developed, data were tracked in an online form. The web app went live in August 2021, and all records prior to that point were batch imported into the web app platform. Between April 2021 and March 2022, 439 unique patients completed 5087 forms using the web app. A majority of the forms were from daily monitoring (4430/5087, 87%), followed by screening (439/5087, 9%) and discharge (218/5087, 4%) forms. A total of 14 staff members were initially trained on how to use the web app and complete the forms. The staff received informal support from the technical team through tutorials on logging in, resetting passwords, and using the web app. Paper forms were used to collect data as needed, such as when computers or support staff became unavailable. The paper forms were later entered into the web app database. The software worked as expected, and no critical bugs or issues were reported.

Discussion

Principal Findings

Our international, multidisciplinary partnership successfully developed and implemented a free, open-source web app to remotely monitor COVID-19 patients in Armenia at home by leveraging HCD and an agile software development methodology. Developing software with international colleagues required attention to logistics and strategy. Active involvement with local stakeholders in Armenia was a priority. Due to substantial differences in time zones, virtual meetings were held early in the morning in the United States (Pacific time), corresponding to evening in Armenia. Additionally, program administrators from the United States visited colleagues in Armenia to establish rapport and facilitate stakeholder buy-in for the project when travel restrictions were relaxed. We explicitly used an HCD approach to ensure that the web app was linguistically and culturally relevant and responsive to the local technical environment and workflows. An important requirement for adapting the software for the end user was the integration of language translations for all components of the web app. We contracted the software development to a local team in Armenia and communicated with local stakeholders, which allowed us to test and iterate the web app more effectively.

We searched PubMed for publications describing COVID-19 patients who were discharged home with supplemental oxygen therapy and their remote monitoring and management using mobile or web apps with the following key words and Medical Subject Headings (MeSH) terms: (COVID-19/therapy[Mesh]) AND (Telemedicine[Mesh] OR Home Care Services[Mesh] OR Telemetry[Mesh] OR Patient Discharge[Mesh] OR telemedicine OR home OR telemetry OR remote monitor* OR virtual) AND (Oxygen Inhalation Therapy[Mesh] OR oxygen). We identified a total of 5 studies, which took place in the Netherlands (n=1), United States (n=2), United Kingdom (n=1), and Egypt (n=1) [28-32]. Aside from the Egyptian study, all the studies took place in high-income countries (HICs). This imbalance reflects the technical, infrastructural, and financial health care divide between HICs and LMICs. All 5 studies had 2 common features: collection of vital signs and communication via telephone [28-32]. Progress assessment and vital signs, such as oxygen saturation, heart rate, temperature, and blood pressure, were captured by all the studies and then stored in electronic databases, not unlike our current project. Capturing health data in electronic databases helps support research, collaboration, and quality improvement [33]. Although all the studies involved a mobile app (for education, assessment, or monitoring vital signs), they all relied on communication with the clinical team via phone calls [28-32]. Our program operated similarly, which supports the utility of leveraging standard forms of communication for remote monitoring. Three of the 5 studies conducted patient satisfaction surveys, with all of them reporting 94% or greater satisfaction with the remote monitoring program and emphasizing the benefit and utility to patients [28,29,32]. With respect to the development of these programs, only 1 described the process in detail [30]. The Atrium Health Hospital at Home (AH-HaH) program was designed and implemented rapidly in the United States with the help of numerous stakeholders with backgrounds including clinical medicine, administration, research, technology, and innovation [30]. This mirrors our process and reiterates the multidisciplinary nature of developing home-based clinical programs supported by technology. However, the AH-HaH program was implemented using existing infrastructure and resources that are not often available in LMICs [30]. Evidently, our program in Armenia was developed with a limited amount of existing infrastructure and operational resources. Other remote monitoring programs have also commented on the importance of a flexible and nimble development framework due to changes in COVID-19 guidelines, an observation that reflected our experience in Armenia as well [34].

The main strength of our project was the rapid, coordinated, and collaborative approach. With logistical uncertainty, limited knowledge, and the evolution of COVID-19 and its variants, it was important to quickly and responsively develop health screening and monitoring forms with the most recent available knowledge. This required updating form questions, the criteria, and form fields as more evidence became available. In turn, the software development cycle reflected this process and required updating code and functionality simultaneously. The conflict with Azerbaijan in 2020 had further strained the Armenian health care system, so the iterative, agile methodology allowed the team to be flexible and responsive to evolving capacity. Other strengths were the reliance on scientific evidence and the involvement of clinicians in the development process. Although case studies and literature on COVID-19 were available, the novelty of the disease and evolution of new strains led to uncertainty. Knowledge of the disease within the context of Armenia was even more limited. However, frontline workers in Armenia helped design the clinical protocols and data collection forms and were informed by their experiences and the best available evidence. This project was truly a collaborative effort, involving individuals of diverse professional experiences, cultures, disciplines, and knowledge. Each feature and change was tested by stakeholders and end users, who provided guidance on adapting the system for local use, such as suggesting language changes and workflow updates. The

COVID@Home program would not have been possible without a multidisciplinary and collaborative team of clinicians, software developers, researchers, and local stakeholders. Finally, COVID@Home set a precedent for active data collection, reporting, monitoring, and outcome evaluations outside of a research setting, which is rarely done in Armenia.

A significant limitation of our approach was the reliance on hard-copy forms, which affected the quality of the data. At various points during the project, computers were only intermittently available for data entry. This resulted in patient data being recorded on paper forms instead of the web app. By the time this was brought to the attention of the project team, hundreds of patient records had been created using paper forms. Dedicated computers were provided for data entry, web app access, and digitizing the paper forms, but this resulted in several records with nonstandardized or missing data. During this time, the software continued to operate as expected. A general challenge across many global health partnerships is the long-term sustainability of a project [35]. This is a critical component of global health partnerships and is built into all of our collaborations in Armenia. For this particular project, the total number of patients enrolled in the program has drastically decreased since July 2022 as COVID-19 hospitalizations have decreased, and the plan from the Ministry of Health is to sunset this program. The technical infrastructure is maintained by a local team that supports multiple grant-funded projects, and they are committed to continue operating the existing system. As this was the first time a home health-monitoring program was used by the Ministry of Health, the ministry is currently evaluating whether to expand this approach to other use cases.

A key lesson learned from this experience was the importance of clear communication and training for all users, particularly the frontline HCPs. A more thorough needs and capacity assessment likely would have detected the issue of paper-based data collection; although the project team asked if a computer was available, additional follow-up questions may have revealed that there were limitations affecting who could access the computer, how often, and how often there were interruptions to that access. Prospective utilization monitoring might also have detected that new records were not being created, prompting further investigation. As an operational initiative leveraging existing staff and resources, the project did not operate in a vacuum, and was affected by broader factors in the health system: lack of human resources, lack of planning, and lack of proper equipment. For example, in one instance, the project coordinator was reassigned, leaving frontline staff with no support for several weeks. The development of paper forms played an important and pragmatic role in the success of the program, though once staff transitioned to paper forms due to computer access issues, it was difficult to shift the workflow back to the web app. In a less chaotic environment, not having paper forms at all may help increase the speed of adoption and prevent parallel workflows. Future projects will need to better evaluate cultural, sociotechnological, and logistical barriers to technology adoption in busy health care settings.

Conclusion

COVID-19 has caused tremendous clinical and logistical challenges around the globe, particularly in LMICs. Our experience in Armenia demonstrates that an international collaboration can leverage an agile methodology to rapidly develop software needed to support a home monitoring program for patients with prolonged oxygen requirements secondary to COVID-19 using free, open-source software. This project can serve as an example for at-home care in resource-limited settings like Armenia and other LMICs. Successful implementation requires a thorough needs assessment, clear and open communication, and prospective utilization monitoring to identify and address barriers.

Acknowledgments

The authors would like to thank Arman Veranyan and the software development team in Armenia for their tremendous effort in developing the technology. We also thank our collaborators from the Armenian EyeCare Project and the University of California, Los Angeles Project Armenia for their dedication and support in bringing this project to fruition. This project was supported by an anonymous gift. JE's time was supported in part by grants from the National Center for Advancing Translational Science (UL1TR001855 and UL1TR000130) of the US National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

JE is a paid consultant for AI Health. AI Health had no role in this work. None of the other authors have any disclosures to report.

References

1. Coronavirus (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2022-11-15]
2. Afewerky HK. Pathology and pathogenicity of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Exp Biol Med (Maywood) 2020 Sep;245(15):1299-1307 [FREE Full text] [doi: 10.1177/1535370220942126] [Medline: 32635753]
3. Yesudhas D, Srivastava A, Gromiha MM. COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. Infection 2021 Apr;49(2):199-213 [FREE Full text] [doi: 10.1007/s15010-020-01516-2] [Medline: 32886331]
4. Ejaz H, Alsrhani A, Zafar A, Javed H, Junaid K, Abdalla AE, et al. COVID-19 and comorbidities: Deleterious impact on infected patients. J Infect Public Health 2020 Dec;13(12):1833-1839 [FREE Full text] [doi: 10.1016/j.jiph.2020.07.014] [Medline: 32788073]

5. Maves RC, Downar J, Dichter JR, Hick JL, Devereaux A, Geiling JA, ACCP Task Force for Mass Critical Care. Triage of scarce critical care resources in COVID-19 an implementation guide for regional allocation: an expert panel report of the task force for mass critical care and the American College of Chest Physicians. Chest 2020 Jul;158(1):212-225 [FREE Full text] [doi: 10.1016/j.chest.2020.03.063] [Medline: 32289312]
6. Shahzad H, Mubarik F, Sattar AK. The novel coronavirus (COVID-19) pandemic and the response in low-to-middle income countries. Curr Breast Cancer Rep 2021;13(2):63-68 [FREE Full text] [doi: 10.1007/s12609-021-00409-z] [Medline: 33880139]
7. General information about Republic of Armenia. Ministry of Foreign Affairs of the Republic of Armenia. URL: https://www.mfa.am/en/overview [accessed 2022-11-15]
8. Kazaryan AM, Edwin B, Darzi A, Tamamyan GN, Sahakyan MA, Aghayan DL, Fretland, Doctors Against The War collaborators. War in the time of COVID-19: humanitarian catastrophe in Nagorno-Karabakh and Armenia. Lancet Glob Health 2021 Mar;9(3):e243-e244 [FREE Full text] [doi: 10.1016/S2214-109X(20)30510-6] [Medline: 33253632]
9. A Balalian A, Berberian A, Chiloyan A, DerSarkissian M, Khachadourian V, Siegel EL, Public Health Working Group for Armenia. War in Nagorno-Karabakh highlights the vulnerability of displaced populations to COVID-19. J Epidemiol Community Health 2021 Mar 05:jech-2020-216370. [doi: 10.1136/jech-2020-216370] [Medline: 33674457]
10. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, et al. Coronavirus (COVID-19) Cases. Our World in Data. URL: https://ourworldindata.org/covid-cases [accessed 2022-11-15]
11. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2022-11-16]
12. Sikder A, Mehta P, Galoyan T, Parekh E, Dickhoner J, Yeghiazaryan N, et al. Using a Facebook virtual examination room to enhance ROP training in Armenia. Clin Ophthalmol 2022;16:2311-2322 [FREE Full text] [doi: 10.2147/OPTH.S371360] [Medline: 35911845]
13. Jrbashyan N, Yeghiazaryan N, Sikder A, Nallasamy S, Reid MW, Ohanesian R, et al. Pattern and prevalence of eye disorders and diseases in school-aged children: findings from the Nationwide School Sight Sampling Survey in Armenia. BMJ Open Ophthalmol 2022;7(1):e000899 [FREE Full text] [doi: 10.1136/bmjophth-2021-000899] [Medline: 35265749]
14. Galoyan T, Kysh L, Lulejian A, Dickhoner J, Sikder A, Lee M, et al. Lessons learned from organizing and evaluating international virtual training for healthcare professionals. Int J Med Educ 2022 Mar 31;13:88-89 [FREE Full text] [doi: 10.5116/ijme.6238.459f] [Medline: 35367957]
15. Espinoza J, Sikder AT, Dickhoner J, Lee T. Assessing health data security risks in global health partnerships: development of a conceptual framework. JMIR Form Res 2021 Dec 08;5(12):e25833 [FREE Full text] [doi: 10.2196/25833] [Medline: 34889752]
16. Telementoring: How Technology is Helping to Lower the Impact of ROP in Armenia. Retina Today. URL: https://retinatoday.com/articles/2015-may-june/telementoring-how-technology-is-helping-to-lower-the-impact-of-rop-in-armenia [accessed 2022-11-15]
17. Operation Armenia: The Promise Armenian Institute. University of California, Los Angeles. URL: https://www.international.ucla.edu/armenia/operationarmenia [accessed 2022-11-15]
18. Introduction to OpenMRS - Documentation - OpenMRS Wiki. OpenMRS. URL: https://wiki.openmrs.org/display/docs/Introduction+to+OpenMRS [accessed 2022-11-15]
19. OpenMRS Response to COVID-19. OpenMRS. URL: https://openmrs.org/2020/03/11/openmrs-response-to-covid-19/ [accessed 2022-11-15]
20. Lim HM, Teo CH, Ng CJ, Chiew TK, Ng WL, Abdullah A, et al. An automated patient self-monitoring system to reduce health care system burden during the COVID-19 pandemic in Malaysia: development and implementation study. JMIR Med Inform 2021 Feb 26;9(2):e23427 [FREE Full text] [doi: 10.2196/23427] [Medline: 33600345]
21. Basit MA, Baldwin KL, Kannan V, Flahaven EL, Parks CJ, Ott JM, et al. Agile acceptance test-driven development of clinical decision support advisories: feasibility of using open source software. JMIR Med Inform 2018 Apr 13;6(2):e23 [FREE Full text] [doi: 10.2196/medinform.9679] [Medline: 29653922]
22. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
23. Data Model - Documentation - OpenMRS Wiki. OpenMRS. URL: https://wiki.openmrs.org/display/docs/Data+Model [accessed 2022-11-15]
24. Getting and Using the CIEL Concept Dictionary - Documentation - OpenMRS Wiki. OpenMRS. URL: https://wiki.openmrs.org/pages/viewpage.action?pageId=34148035 [accessed 2022-11-15]
25. Spring | Home. VMware. URL: https://spring.io/ [accessed 2022-11-15]
26. Technical Overview - Documentation - OpenMRS Wiki. OpenMRS. URL: https://wiki.openmrs.org/display/docs/Technical+Overview [accessed 2022-11-15]
27. Everything data. Hibernate. URL: https://hibernate.org/ [accessed 2022-11-15]

28. Grutters LA, Majoor KI, Mattern ESK, Hardeman JA, van Swol CFP, Vorselaars ADM. Home telemonitoring makes early hospital discharge of COVID-19 patients possible. J Am Med Inform Assoc 2020 Nov 01;27(11):1825-1827 [FREE Full text] [doi: 10.1093/jamia/ocaa168] [Medline: 32667985]
29. Huynh DN, Millan A, Quijada E, John D, Khan S, Funahashi T. Description and early results of the Kaiser Permanente Southern California COVID-19 home monitoring program. Perm J 2021 Jul 28;25:1-7 [FREE Full text] [doi: 10.7812/TPP/20.281] [Medline: 35348067]
30. Sitammagari K, Murphy S, Kowalkowski M, Chou S, Sullivan M, Taylor S, et al. Insights from rapid deployment of a "virtual hospital" as standard care during the COVID-19 pandemic. Ann Intern Med 2021 Feb;174(2):192-199 [FREE Full text] [doi: 10.7326/M20-4076] [Medline: 33175567]
31. O'Malley E, Hansjee S, Abdel-Hadi B, Kendrick E, Lok S. A Covid -19 virtual ward model: a preliminary retrospective clinical evaluation from a UK district general hospital. J Prim Care Community Health 2022;13:21501319211066667 [FREE Full text] [doi: 10.1177/21501319211066667] [Medline: 34986693]
32. Adly AS, Adly MS, Adly AS. Telemanagement of home-isolated COVID-19 patients using oxygen therapy with noninvasive positive pressure ventilation and physical therapy techniques: randomized clinical trial. J Med Internet Res 2021 Apr 28;23(4):e23446 [FREE Full text] [doi: 10.2196/23446] [Medline: 33819166]
33. Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low- and middle-income countries. Bull World Health Organ 2015 Mar 01;93(3):203-208 [FREE Full text] [doi: 10.2471/BLT.14.139022] [Medline: 25767300]
34. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson JA, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]
35. Binagwaho A, Allotey P, Sangano E, Ekström AM, Martin K. A call to action to reform academic global health partnerships. BMJ 2021 Nov 01;375:n2658. [doi: 10.1136/bmj.n2658] [Medline: 34725093]

Abbreviations

AH-HaH: Atrium Health Hospital at Home
API: application programming interface
CHLA: Children's Hospital Los Angeles
CSV: comma-separated value
DL: database layer
EHR: electronic health record
HCD: human-centered design
HCP: home care provider
HIC: high-income countries
LMIC: low- and middle-income country
MeSH: Medical Subject Headings
PL: presentation layer
SL: service layer
UAT: user acceptance testing
UCLA: University of California, Los Angeles

Edited by A Kushniruk; submitted 20.07.22; peer-reviewed by A Bazzano, J Martin; comments to author 06.10.22; revised version received 14.10.22; accepted 07.11.22; published 25.11.22.

Please cite as:
Sikder A, Dickhoner J, Kysh L, Musheghyan L, Shekherdimian S, Levine B, Espinoza J
International Collaboration to Develop a Remote Monitoring Web App for COVID-19 Patients in Armenia: Design and Development With Agile Methodology
JMIR Hum Factors 2022;9(4):e40110
URL: https://humanfactors.jmir.org/2022/4/e40110
doi:10.2196/40110
PMID:36350739

Original Paper

Usability Evaluation of the Preoperative ISBAR (Identification, Situation, Background, Assessment, and Recommendation) Desktop Virtual Reality Application: Qualitative Observational Study

Eva Mari Andreasen[1*], MD; Rune Høigaard[2*], PhD; Helen Berg[3*], PhD; Aslak Steinsbekk[4*], PhD; Kristin Haraldstad[1*], PhD

[1]Department of Health and Nursing Sciences, University of Agder, Kristiansand, Norway
[2]Department of Sport Science and Physical Education, University of Agder, Kristiansand, Norway
[3]Department of Health Sciences, Norwegian University of Science and Technology, Ålesund, Norway
[4]Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway
[*]all authors contributed equally

Corresponding Author:
Eva Mari Andreasen, MD
Department of Health and Nursing Sciences
University of Agder
P.O. Box 422
Kristiansand, 4604
Norway
Phone: 47 90642121
Email: eva.mari.andreasen@uia.no

Abstract

Background: Systematic communication, such as the ISBAR (identification, situation, background, assessment, recommendation) approach, comprises a generic, transferable nontechnical skill. It can be used during the handover of patients set to undergo surgery and can be practiced in various ways, including virtual reality (VR). VR increasingly has been implemented and valued in nursing education as a positive contribution to teach students about pre- and postoperative nursing. A new nonimmersive 3D learning activity called the Preoperative ISBAR Desktop VR Application has been developed for undergraduate nursing students to learn preoperative handover using the ISBAR approach. However, the usability of this learning activity has not been studied.

Objective: This study aimed to investigate how second-year undergraduate nursing students evaluated the usability of the Preoperative ISBAR Desktop VR Application.

Methods: This was a qualitative study with observation and interviews. The inclusion criteria were undergraduate second-year nursing students of varying ages, gender, and anticipated technological competence. The System Usability Scale (SUS) questionnaire was used to get a score on overall usability.

Results: A total of 9 second-year nursing students aged 22-29 years participated in the study. The average score on the SUS was 83 (range 0-100), which equals a “B” on the graded scale and is excellent for an adjective-grade rating. The students expressed increased motivation to learn while working in self-instructed desktop VR. Still, a few technical difficulties occurred, and some students reported that they experienced some problems comprehending the instructions provided in the application. Long written instructions and a lack of self-pacing built into the application were considered limitations.

Conclusions: The nursing students found the application to be usable overall, giving it an excellent usability score and noting that the application provided opportunities for active participation, which was motivational and facilitated their perceived learning outcomes. The next version of the application, to be used in a randomized controlled trial, will be upgraded to address technological and comprehension issues.

(JMIR Hum Factors 2022;9(4):e40400) doi:10.2196/40400

KEYWORDS

desktop virtual reality; handover; ISBAR; preoperative; undergraduate nursing students; usability evaluation; usability; nursing; health care education; student; medical education; medical training; VR; virtual reality; surgery; surgical; System Usability Scale; communication; self-instruction; self-guided; nurse; training; undergraduate; health care professional; health care provider

Introduction

The exchange of relevant clinical information from one provider to another (eg, handover) is crucial for the surgical pathway because missing information and incomplete handover can lead to adverse patient outcomes [1,2]. The ISBAR (identification, situation, background, assessment, and recommendation) approach is an evidence-based approach to ensure consistent, structured communication [3] and can be used in inter- and intraprofessional collaboration for patients about to undergo surgery [4-8]. Studies have reported that using ISBAR can improve communication between health care providers [9,10] and reduce communication errors [11].

Considering that a lack of clear communication directly or indirectly can endanger patient safety, the evidence suggests that ISBAR skills acquisition should start early in nursing education [12,13]. ISBAR traditionally is learned through role-playing in simulations or in classroom settings [14,15]. The past few years have seen an increased interest in virtual reality (VR) as a method to learn structured communication [16-18].

Desktop VR is a computer-generated 3D environment presented on nonimmersive desktop and laptop PC screens [19]. Desktop VR typically is built around user interaction, such as moving avatars, typing commands, and interacting with others while completing a task [19]. The advantage of desktop VR is that it has potential for letting users practice without supervision while receiving audio and visual instruction and has instant feedback from the VR application itself in a safe environment [17]. VR has been increasingly implemented and valued in nursing education as a positive contribution to curricula to teach students about pre- and postoperative nursing [20-22]. Using desktop VR as an active learning method also aligns with studies that have recommended interactive teaching strategies in curricula [23]. However, to the best of our knowledge, no published research exists on desktop VR solutions that practice handover using the ISBAR approach in a preoperative setting [24].

Perceived usability is essential when developing such solutions [25-27]. The International Organization for Standardization has defined *usability* as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [28]. Furthermore, the degree of learnability is defined as part of the usability assessment [28]. We developed the Preoperative ISBAR Desktop VR Application (henceforth "application"), which is intended to be used in a randomized controlled trial. Thus, its usability needs to be tested to optimize the application for virtual simulation in nursing education.

This study aimed to investigate how second-year undergraduate nursing students evaluated the application's usability.

Methods

Design

This was a qualitative observational study with interviews. The usability test was conducted during the fall semester of 2021.

Preoperative ISBAR Desktop VR application

The application was part of a research project focusing on the use of VR in health care education [29]. It was created to teach handover skills when using the ISBAR approach and is based on cognitive principles from the 4-component instructional design (4C/ID) [30] guidelines, comprising (1) a learning task, (2) supportive information, (3) procedural information, and (4) part-task practice. Instructions and tasks were based on evidence-based knowledge of learner-centered teaching [31-33] and national ISBAR guidelines [34]. A version of the application still under development was used.

The various sequences in the application are presented in Table 1. The students were organized into groups of 3 who played together in VR through 3 main activities. The first activity was to sort patient information individually using the ISBAR approach and compare and discuss the participants' individual sorting to clarify the ISBAR approach. The second activity was to perform handovers using the ISBAR approach, which was between a nurse on a night shift and a day shift, and between a nurse on a day shift and an anesthesia nurse. The third activity was a debriefing that focused on the experience in general and on selecting the most important patient information to communicate first.

Table 1. Presentation of the Preoperative ISBAR[a] Desktop VR Application.

Number	Sequence	Content
1	Instruction: register name and select group number	A screen with a visible square to insert the participant's name. Group allocation number visible with instruction to choose groups
2	Instruction: introduction to ISBAR	Animation with a voiceover explaining briefly what ISBAR is, presenting the learning objectives, and providing a brief overview of the tasks
3	Task: familiarization with desktop VR[b] and each other	A screen displays instructions on how to use the arrow keys to look around in the desktop VR and introduce the players to each other
4	Instruction: sort patient information	Animation with a voiceover instructing how to sort single pieces of patient information according to the ISBAR approach and how to get additional information in pop-up windows
5	Task: sort patient information	A screen with an area displaying 1 piece of patient information, buttons for each ISBAR letter to select where the patient information should be sorted, and a list containing patient information sorted in the order of the selected ISBAR letters with the opportunity to delete the patient information and to sort it again. An explanation of the ISBAR approach is available as a pop-up
6	Task: discussion of experience with sorting	A screen displays a comparison of how each participant sorted the patient information and a suggestion for correct sorting. The percentage of patient information sorted similarly to the suggested solution is displayed for each player
7	Instruction and task: patient case and choose a role	Animation with a voiceover presenting a patient case, the 3 roles involved (nurse on night shift, nurse on day shift, and nurse anesthetist), and how to choose a role
8	Instruction and task: role description and choose a role	A screen with a written description of the 3 roles involved and pictures symbolizing the roles to be clicked to select a role. When a player clicks on a role, the frame changes to green for that player and red for the other players
9	Instruction: handover role play	Animation with a voiceover instructing how to complete the next task and a handover role play in which participants give and receive patient information in their active roles (nurse on night shift, nurse on day shift, and nurse anesthetist) using the ISBAR approach. Instruction on active participation for both the giver and receiver in handover (sender starts with the selected patient information, and the receiver requests additional patient information)
10	Instruction: handover role play	A screen displays a written summary of the next task
11	Task: handover role play	A screen displays a list of all patient information and a virtual phone. The text states that the player should select the patient information to be presented first and then call the next nurse to perform the handover through the virtual phone. The phone and handover checklist are visible to the receiver of the handover. An explanation of the role play is available as a pop-up for all players during handover practice. The participant's screen with the active role is visible to the other participants in the group. ISBAR explanation available as a pop-up
12	Instruction: debriefing 1	Short animation with a voiceover describing what should be done during the debriefing session
13	Task: debriefing 1	Text stating that they should discuss each participant's experience doing the tasks in general and that they will discuss each participant in detail afterward
14	Instruction: debriefing 2	Animation with a voiceover with instructions to debrief what each participant chose to highlight and say first during the handover
15	Task: debriefing 2	A screen displays a list of all patient information, with the patient information that the participant had clicked on as the information to present first in bold (highlighted). Suggested bullet points on what to discuss during the debriefing are visible as a pop-up explanation. An ISBAR explanation is available as a pop-up
16	Instruction: debriefing closure	Animation with a voiceover with encouragement to practice again
17	Task: final practice and ending	A screen with available options: to practice again or end the practice. If selecting to practice again, it starts at sequence 2

[a]ISBAR: identification, situation, background, assessment, and recommendation.

[b]VR: virtual reality.

Participants and Recruitment

The aim was to include undergraduate second-year nursing students with variations in age, gender, and anticipated technological competence. With 3 students participating in each group, 9 participants were viewed as adequate for robust usability to get a measure of the perceived usability and to get a good assessment of how people see a system or a product [35].

Information about the study was presented verbally in a compulsory lecture for the second-year nursing students at a university in Norway. Furthermore, written study information and recruitment invitations were displayed on a web-based notice board. Those interested were asked to contact the study, and if they did, they received more detailed information about the study, and an appointment for a test time was set. The students were assigned to the 3 groups based on the order in which they signed up for the study.

Procedure and Data Collection

Overview

The whole learning activity comprised watching a 9-minute video introducing ISBAR [36] and practicing within the application. Three students in each group were placed in separate rooms to ensure that all communication happened in the application, mimicking a situation in which the students were in different locations. One researcher was present in each room to observe and provide support if needed.

Data were collected through (1) background questions, (2) observation, (3) the System Usability Scale (SUS), and (4) focus group interviews.

Background Questions

The participants were asked about their gender, age, and whether they had participated in compulsory ISBAR teaching (yes/no). The participants were also asked about their self-reported technological competence, measured on a 4-point graded scale developed for this study, ranging from level 1 (*low competence*) to level 4 (*high competence*).

Observation

The students were encouraged to think aloud, that is, verbalize their thoughts, constantly [37] while using the application. The think-aloud sessions were video-recorded, and field notes were taken based on a predefined observation template covering navigation errors, ease of use, apparent misunderstandings, and technical difficulties (Multimedia Appendix 1). If the students were unsure of how to proceed with the application, they were encouraged to do what they would find most intuitive before being assisted, as Rubin and Chisnell [25] recommended.

System Usability Scale

The participants were asked to complete the SUS [38] after they finished using the application. SUS is a recommended tool for evaluating educational technology systems [26], comprising 10 open-ended items with 5 answer options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The final mean score ranges from 0 to 100, and the score can be reported as an A-F grade using a curved grading scale [39], and as an adjective score, ranging from *worst imaginable* to *best imaginable* [40].

Focus Group Interviews

After completing the SUS, a focus group interview was conducted with each group. An interview guide—which was developed based on the research question, predefined observation template, and usability theory [25]—was used (Multimedia Appendix 2). Examples of questions asked included, “What did you like the most about learning Preoperative ISBAR in desktop VR?,” “What did you like the least about learning Preoperative ISBAR in desktop VR?,” and “Was there anything that exhausted you during the learning activity? If so, what caused the exhaustion?” Furthermore, the interviews addressed specific usability issues observed when the participants completed the application. Each interview lasted approximately 35-40 minutes, and the interview sessions were audio-recorded.

Analysis

The data were analyzed using different approaches. The average score from the SUS questionnaire was calculated using the procedure described by Brooke [38], presented as mean and SD values, and then given a graded score (A-F) based on the acceptability range. The average adjective score was calculated as recommended [40]. Data on task completion time (efficiency) were gathered from field notes and video recordings and presented with descriptive statistics.

All material (video recordings, field notes from the think-aloud sessions, and transcribed focus group interviews) was analyzed together as recommended by Rubin and Chisnell [25], for completeness and to obtain an overview during analysis. The first author transcribed all audio-recorded material (think-aloud sessions and focus group interviews). The transcribed material was analyzed with the field notes, as recommended by Rubin and Chisnell [25], for completeness and to obtain an overview during analysis. A reflexive thematic analysis [41] was conducted to identify in-depth usability issues, with an emphasis on participants’ experiences. The first author led the analysis of the audio-recorded material and field notes to ensure consistency, but the coauthors reviewed and discussed the analysis until an agreement between the coauthors and the first author was reached.

Ethical Considerations

Permission was obtained from the head of the nursing study program at the Department of Health and Nursing Sciences at the University of Agder, the Faculty Ethics Committee at the University of Agder, and the Norwegian Center for Research Data (305866).

Results

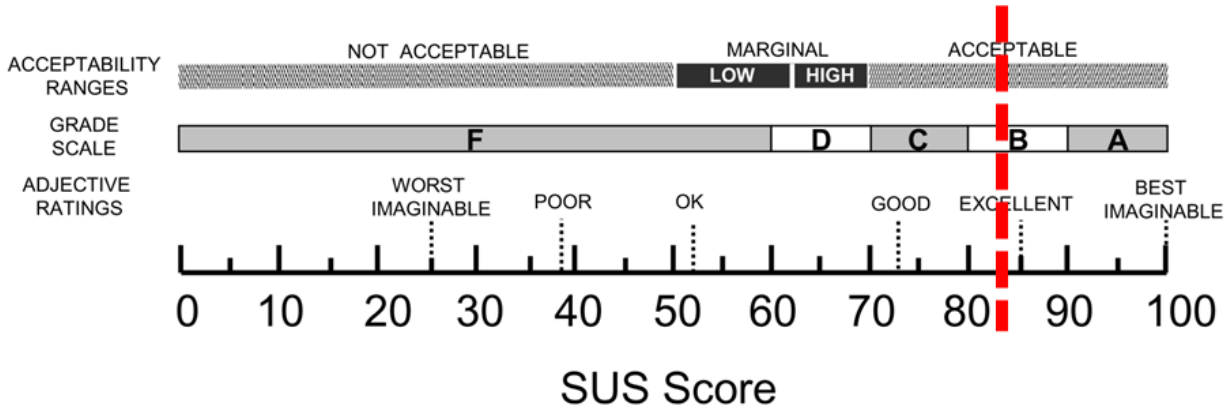
Participants

A total of 9 students responded, and all were included, comprising 7 females and 2 males ranging in ages between 22 and 29 years. All participants previously had taken part in compulsory ISBAR teaching in nursing education. The participants reported their technological competency to be either level 2 (n=5) or level 3 (n=4).

Overall System Usability Assessment

The overall mean SUS score for the application was 83 (SD 18.8; Figure 1), which rates as *Acceptable* on the acceptability range, *B* on the graded scale, and *Excellent* on the adjective rating scale (Figure 1).

Figure 1. Overall system usability assessment. The vertical dotted red line (83 on the 0-100 scale) shows the mean system usability score (SUS) (n=9) (reproduced with permission from Aaron Bangor [42]). SUS: System Usability Scale.



Time

The 3 groups took 28, 37, and 48 minutes to complete all sequences in the application once, with a mean time of 38 minutes. The group that took 48 minutes had one participant who spent 15 minutes sorting patient information (Table 1, sequence 5), while the other participants used 3 to 5 minutes on this task.

In-depth Usability Issues (Thematic Analysis)

The qualitative findings were categorized into two themes: (1) more motivational than standard learning activities and (2) technical and comprehension issues.

More Motivational Than Standard Learning Activities

All participants recommended this learning activity to others and said it was a motivational way of learning ISBAR. The participants said the application helped them learn ISBAR through the self-instructed exercise, discussions with other participants, observing how others performed the tasks, and when the instructions told them to reflect on their performance together with the others.

> *It was good to practice ISBAR instead of just reading. It is actually better to do it. It is more like reality and a lot more fun. Communication is a skill. By reading about ISBAR, you will never be good at communicating. It is a skill that must be practiced. By using this, you are practicing communication. You can memorize the letters in ISBAR, but you cannot use it if you haven't practiced.* [ID 07, self-reported technological competence level 2]

All the participants concluded that the application's features, such as the automatic visualized feedback, motivated them to complete the exercise. Some of them said that being represented as avatars with their own voices and not having to reveal themselves on camera was a good way to practice. Furthermore, some commented that communication through the virtual phone call made them realize that they needed to speak clearly and loudly.

> *I feel that I am more invested in it because it is a PC program. It could be a desire for learning or a competitive instinct, but I want to complete this program. I liked the feeling of progression and the structure. Everyone knew what we were going to do, and we knew what to do. It is systematic, and you go on and on and on and on.* [ID 02, self-reported technological competence level 3]

Some also described it as being closer to clinical practice compared with standard learning activities.

> *It was simulated in a way that made me feel that I got something out of it, and it was a good way to go from theory to practice. It is like a clinical procedure; you do not know how to do it by reading the procedure, but by doing it. You learn to use ISBAR during clinical practice, but having a good start like this can make you learn faster and better.* [ID 01, self-reported technological competence level 3]

Technical and Comprehension Issues

All groups managed to complete the application and all the users were able to complete the tasks "Familiarization with desktop VR and each other" and "Sort patient information" at the first attempt. Screen transitions were crisp and smooth, with no apparent technical lag times that may have led to negative usability. Through the interviews, most participants said it was easy to follow the application's flow and complete the tasks.

> *Technically, it is very easy to understand. For me, it looks like anyone could have managed this. If a technical manager had assisted, that person would not have needed to help them much. It was obvious.* [ID 05, self-reported technological competence level 3]

However, in 2 of 3 usability tests, the application was restarted. In 1 test, a participant had trouble getting access to the microphone on the computer, so the other participants had to wait until this was solved. During another test, a participant clicked on the "Next" button on the screen at a point when they

are supposed to introduce themselves to each other (Table 1, sequence 3). Thus, they were instructed to restart the application. When these 2 issues were resolved, the groups completed the tasks.

> *The "technical" problems that occurred were frustrating. I was terrified of doing something wrong. I understood that if you clicked "Next," everyone must start again. So, I got stressed because of the disturbances initially when we had to start over again. Then I thought that I would not ruin it for everyone else. And then I just got even more stressed.* [ID 04, Self-reported technological competence level 2]

Some participants said they had problems understanding parts of each task. One reason was that they did not hear the voiceover instructions (Table 1, sequences 2, 4, 7, and 9) owing to other participants commenting or asking questions during the instructions. Another reason given was reluctance to open the available pop-up windows to repeat the instructions for fear of appearing slow or incompetent to other participants. Finally, some said that the most prolonged instructions contained too much information (Table 1, sequences 8 and 10), making them forget what was said.

> *I did not really understand whether we should include everything or not. That was the hardest to understand. I think it was because I did not read the instructions before. I was stressed, feeling the others may read faster than me. And I am slow, so I just had to hurry, right? And then I did not read the instructions.* [ID 09, self-reported technological competence level 2]

During task completion, 2 of the 9 participants asked for instructions from the observer in the room. The requested instructions were in sequence 5 (Table 1), when it was unclear whether they should answer individually or in a group, and in sequence 11, when someone asked for instructions on how to solve the task regarding whether they should sort all patient information or only some of it.

Discussion

Principal Findings

This study aimed to identify the perceived usability of the application as evaluated by second-year nursing students, who found the learning activity to be usable overall, rating it highly, although with some technical and comprehension issues that impeded the experience for some testers.

Recommended Changes to the Preoperative ISBAR Desktop VR Application

As described, usability issues were found, and it is recommended that such issues be addressed by making changes to the application. Some participants took an unnecessarily long time to complete some tasks, for example, trying to perfect their answers. However, this may be due to the experimental task given and not a usability issue. Nevertheless, it is recommended to impose a time limit for some tasks (Table 1, sequences 5 and 11, with a time limit of 5.5 minutes and 1 minute, respectively) to more accurately reflect the practical context (eg, time pressure, stress, and workload).

Considering that participants were disturbed when other participants talked during the instructions, it is recommended that participants be muted while instructions are given (Table 1, sequences 2, 4, 7, and 9). To avoid the participant clicking on the "Next" button too early, a 10-second delay after the spoken instructions are completed before "Next" can be clicked is recommended. Furthermore, it is recommended that each written instruction sequence and task sequence start with the informative pop-up windows open so that they only can be closed manually to allow all participants to read or reread through the information at their own pace, which is an appealing approach for students [43].

VR as a Learning Technology

All participants found the VR application to be a motivational way of learning ISBAR. Using desktop VR for learning purposes seems to fit the targeted users, which is perhaps not surprising, as they all were born in the mid-1990s as part of a demographic termed Generation Z [44]. This generation grew up with access to the internet and digital technology from a young age [45]. According to Chicca and Shellenbarger [45], this generation is supported during the learning process when technologically advanced and visually engaging and exciting activities are provided.

Some students said that the application helped them learn the ISBAR approach better than traditional activities, mainly because they could participate actively and experience the training closer to practice. This supports Huang and Liaw [46], suggesting that a well-designed VR learning environment can bridge the gap between theoretical and real-life learning, providing learners with a more authentic learning experience. The results indicate the application's utility, providing self-reported improvement in the performance of the ISBAR approach compared to conventional training, which could be mediated by the interaction experience and the pedagogical support in the application [47]. Thus, the learning outcome must be further studied using a suitable design to measure the learning effect.

Even if the application's evaluation primarily was positive, the participants also reported some challenges due to negative stress when task completion did not progress as intended. Technological usability issues affect the participants' experiences [43,47]. Furthermore, the individual differences in how people react to using VR for learning [46] need to be considered when designing learning activities.

The students stated that they were reluctant to open the pop-up windows for explanations when everyone in the group was watching. Others' influence has been noted in extant research when participants are observed performing tasks, a phenomenon explained by the social facilitation theory [48,49]. The assumption is that others' presence can both promote and hinder one's performance, which also is supported by Strojny et al's [50] investigation of copresence in VR. In an earlier study, it was suggested that self-paced learning be taken into account

during instruction through desktop VR because it generates autonomy [51].

Methodological Strengths and Limitations

This study's strength was that the participants were the intended user group, who varied in age, gender, and self-reported technological competence. This variation can enhance the generalizability of the results [25]. Nevertheless, some caution is needed because the participants were self-recruited, which could mean they were overly positive about VR and technology-based teaching [52].

Although the SUS is a recommended tool for evaluating educational technology systems and is suitable for a small sample size [26], the scale was not developed specifically to evaluate learning activities in desktop VR. Therefore, the think-aloud method and focus group interviews were supporting methods in this study.

Conclusions

The second-year undergraduate nursing students rated the application's usability as excellent and provided opportunities for active participation, which was motivational and facilitated their perceived learning outcomes. The next version of the application, to be used in a randomized controlled trial and further as a part of clinical preparation in nursing education, will include better technological and comprehension support.

Acknowledgments

We would like to give special thanks to the nine undergraduate nursing students who participated in this study. We also are grateful to research assistant Susanne Aune Solås for her valuable help during data collection.

Authors' Contributions

EMA, HB, AS, RH, and KH contributed to the study design. EMA and KH performed the data collection, and EMA, HB, AS, RH, and KH performed the analysis. EMA drafted the manuscript, and HB, AS, RH, and KH contributed during the manuscript development process. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
An avatar represents each participant, and the interaction takes place in the desktop virtual reality (VR).
[PNG File , 757 KB - humanfactors_v9i4e40400_app1.png]

Multimedia Appendix 2
A picture of the screen for the task of sorting patient information. One part of the patient information is shown in the upper left part and is sorted by clicking on one of the ISBAR letters. To the right, the patient information already sorted is displayed.
[PNG File , 1135 KB - humanfactors_v9i4e40400_app2.png]

References

1. Nagpal K, Arora S, Vats A, Wong HW, Sevdalis N, Vincent C, et al. Failures in communication and information transfer across the surgical care pathway: interview study. BMJ Qual Saf 2012 Oct;21(10):843-849 [FREE Full text] [doi: 10.1136/bmjqs-2012-000886] [Medline: 22773891]
2. Johnston MJ, Arora S, King D, Bouras G, Almoudaris AM, Davis R, et al. A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery. Surgery 2015 Apr;157(4):752-763 [FREE Full text] [doi: 10.1016/j.surg.2014.10.017] [Medline: 25794627]
3. Shahid S, Thomas S. Situation, background, assessment, recommendation (SBAR) communication tool for handoff in health care—a narrative review. Saf Health 2018 Jul 28;4(1):1-9. [doi: 10.1186/s40886-018-0073-1]
4. Nilsson U, Gruen R, Myles PS. Postoperative recovery: the importance of the team. Anaesthesia 2020 Jan;75 Suppl 1:e158-e164 [FREE Full text] [doi: 10.1111/anae.14869] [Medline: 31903575]
5. Kitney P, Tam R, Bennett P, Buttigieg D, Bramley D, Wang W. Handover between anaesthetists and post-anaesthetic care unit nursing staff using ISBAR principles: a quality improvement study. J Perioper Nurs 2017 Mar;29(1):30-35. [Medline: 30708404]
6. Kitney P, Tam R, Bramley D, Simons K. Handover using ISBAR principles in two perioperative sites—a quality improvement project. J Perioperat Nurs 2020 Dec 09;33(4):17-22 [FREE Full text] [doi: 10.26550/2209-1092.1094]
7. Amato-Vealey EJ, Barba MP, Vealey RJ. Hand-off communication: a requisite for perioperative patient safety. AORN J 2008 Nov;88(5):763-774; quiz 771. [doi: 10.1016/j.aorn.2008.07.022] [Medline: 19024783]
8. Leonardsen A, Moen EK, Karlsøen G, Hovland T. A quantitative study on personnel's experiences with patient handovers between the operating room and the postoperative anesthesia care unit before and after the implementation of a structured communication tool. Nurs Rep 2019 Apr 19;9(1):1-5. [doi: 10.4081/nursrep.2019.8041]

9. Raurell-Torredà M, Rascón-Hernán C, Malagón-Aguilera C, Bonmatí-Tomás A, Bosch-Farré C, Gelabert-Vilella S, et al. Effectiveness of a training intervention to improve communication between/awareness of team roles: a randomized clinical trial. J Prof Nurs 2021;37(2):479-487. [doi: 10.1016/j.profnurs.2020.11.003] [Medline: 33867108]
10. Jeong JH, Kim EJ. Development and evaluation of an SBAR-based fall simulation program for nursing students. Asian Nurs Res (Korean Soc Nurs Sci) 2020 May;14(2):114-121 [FREE Full text] [doi: 10.1016/j.anr.2020.04.004] [Medline: 32335316]
11. Randmaa M, Mårtensson G, Swenne CL, Högberg H, Engström M. Implementing situation-background-assessment-recommendation in an anaesthetic clinic and subsequent information retention among receivers: A prospective interventional study of postoperative handovers. Eur J Anaesthesiol 2016 Mar;33(3):172-178.
12. Burgess A, van Diggele C, Roberts C, Mellis C. Teaching clinical handover with ISBAR. BMC Med Educ 2020 Dec 03;20(Suppl 2):459 [FREE Full text] [doi: 10.1186/s12909-020-02285-0] [Medline: 33272274]
13. Gordon M, Hill E, Stojan JN, Daniel M. Educational interventions to improve handover in health care: an updated systematic review. Acad Med 2018 Aug;93(8):1234-1244 [FREE Full text] [doi: 10.1097/ACM.0000000000002236] [Medline: 29620675]
14. Yu M, Kang KJ. Effectiveness of a role-play simulation program involving the sbar technique: a quasi-experimental study. Nurse Educ Today 2017 Jun;53:41-47. [doi: 10.1016/j.nedt.2017.04.002] [Medline: 28433731]
15. Lanz AS, Wood FG. Communicating patient status: comparison of teaching strategies in prelicensure nursing education. Nurse Educ 2018;43(3):162-165. [doi: 10.1097/NNE.0000000000000440] [Medline: 28857952]
16. Liaw SY, Ooi SW, Rusli KDB, Lau TC, Tam WWS, Chua WL. Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. J Med Internet Res 2020 Apr 08;22(4):e17279. [doi: 10.2196/17279] [Medline: 32267235]
17. Shorey S, Ng ED. The use of virtual reality simulation among nursing students and registered nurses: a systematic review. Nurse Educ Today 2021 Mar;98:104662. [doi: 10.1016/j.nedt.2020.104662] [Medline: 33203545]
18. Foronda C, Gattamorta K, Snowden K, Bauman EB. Use of virtual clinical simulation to improve communication skills of baccalaureate nursing students: a pilot study. Nurse Educ Today 2014 Jun;34(6):e53-e57. [doi: 10.1016/j.nedt.2013.10.007] [Medline: 24231637]
19. Ausburn LJ, Ausburn FB. Desktop virtual reality: a powerful new technology for teaching and research in industrial teacher education. J Ind Teach Educ 2004;41(4). [doi: 10.1080/0952398042000314820]
20. Koivisto J, Multisilta J, Niemi H, Katajisto J, Eriksson E. Learning by playing: a cross-sectional descriptive study of nursing students' experiences of learning clinical reasoning. Nurse Educ Today 2016 Oct;45:22-28. [doi: 10.1016/j.nedt.2016.06.009] [Medline: 27429399]
21. Kim H, Lee J, Lee E. Virtual experience of perioperative patients: walking in the patients' shoes using virtual reality and blended learning. Int J Environ Res Public Health 2021 Jun 15;18(12):6457 [FREE Full text] [doi: 10.3390/ijerph18126457] [Medline: 34203661]
22. Tjoflåt I, Brandeggen TK, Strandberg ES, Dyrstad DN, Husebø SE. Norwegian nursing students' evaluation of vSim® for nursing. Adv Simul (Lond) 2018;3:10 [FREE Full text] [doi: 10.1186/s41077-018-0070-9] [Medline: 29946485]
23. Horntvedt MET, Nordsteien A, Fermann T, Severinsson E. Strategies for teaching evidence-based practice in nursing education: a thematic literature review. BMC Med Educ 2018 Jul 28;18(1):172 [FREE Full text] [doi: 10.1186/s12909-018-1278-z] [Medline: 30055612]
24. Andreasen EM, Slettebø Å, Opsal A. Learning activities in bachelor nursing education to learn pre- and postoperative nursing care—a scoping review. Int J Educ Res 2022;115:102033. [doi: 10.1016/j.ijer.2022.102033]
25. Rubin J, Chisnell D. Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests. Indianapolis, IN: John Wiley & Sons; 2008.
26. Vlachogianni P, Tselios N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): a systematic review. J Res Technol Educ 2021 Feb 23;54(3):392-409. [doi: 10.1080/15391523.2020.1867938]
27. Makransky G, Petersen GB. Investigating the process of learning with desktop virtual reality: a structural equation modeling approach. Comput Educ 2019 Jun;134:15-30. [doi: 10.1016/j.compedu.2019.02.002]
28. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. 2016 Presented at: International Conference on Human-Computer Interaction; July 17-22, 2016; The Westin Harbour Castle Hotel, Canada p. 268-278. [doi: 10.1007/978-3-319-39510-4_25]
29. Steinsbekk A. Virtuell Samhandling NTNU Norwegian. VIRSAM. URL: https://www.virsam.no/ [accessed 2022-10-25]
30. Frerejean J, van Merriënboer JJ, Kirschner PA, Roex A, Aertgeerts B, Marcellis M. Designing instruction for complex learning: 4C/ID in higher education. Eur J Educ 2019 Oct 13;54(4):513-524. [doi: 10.1111/ejed.12363]
31. Biggs J, Tang C. Teaching for Quality Learning at University. Maidenhead: McGraw-Hill Education; 2011.
32. Lapum JL, Verkuyl M, Hughes M, Romaniuk D, McCulloch T, Mastrilli P. Self-debriefing in virtual simulation. Nurse Educ 2019;44(6):E6-E8. [doi: 10.1097/NNE.0000000000000639] [Medline: 30585886]
33. Eppich W, Cheng A. Promoting excellence and reflective learning in simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. Simul Healthc 2015 Apr;10(2):106-115. [doi: 10.1097/SIH.0000000000000072] [Medline: 25710312]

34. Tidlig oppdagelse og rask respons ved forverret somatisk tilstand ? nasjonale faglige råd Norwegian. Helsedirektoratet. Oslo: Akershus universitetssykehus: Helsedirektoratet; 2020. URL: https://www.helsedirektoratet.no/faglige-rad/tidlig-oppdagelse-og-rask-respons-ved-forverret-somatisk-tilstand [accessed 2022-10-25]
35. Tullis T, Stetson J. A comparison of questionnaires for assessing website usability. Usability Professional Association Conference. URL: https://www.researchgate.net/publication/228609327_A_Comparison_of_Questionnaires_for_Assessing_Website_Usability [accessed 2022-10-25]
36. Andreasen EM, Haraldstad K, Høigaard R, Berg H, Steinsbekk A. ISBAR. Norway: University of Agder; 2022. URL: https://video.uia.no/media/t/0_yxcdxqlv [accessed 2022-10-25]
37. Charters E. The use of think-aloud methods in qualitative research: an introduction to think-aloud methods. Brock Educ J 2003 Jul 01;12(2):68-82. [doi: 10.26522/brocked.v12i2.38]
38. Brooke J. SUS: a quick and dirty usability scale. Usability Evaluation in Industry. URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2022-10-25]
39. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, CA: Createspace Independent Publishing Platform; 2011.
40. Lewis JR. The System Usability Scale: past, present, and future. Int J Human–Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
41. Braun V, Clarke V. Thematic Analysis: A Practical Guide. London: SAGE; 2022.
42. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usab Stud 2009;4(3):114-123.
43. Kavanagh S, Luxton-Reilly A, Wuensche B, Plimmer B. A systematic review of virtual reality in education. Themes Sci Technol Educ 2017;10(2):85-119.
44. A Dictionary of Human Resource Management. Oxford University Press; 2017. URL: https://www.oxfordreference.com/view/10.1093/acref/9780191827822.001.0001/acref-9780191827822-e-2097 [accessed 2022-10-25]
45. Chicca J, Shellenbarger T. Connecting with generation Z: approaches in nursing education. Teach Learn Nurs 2018 Jul;13(3):180-184. [doi: 10.1016/j.teln.2018.03.008]
46. Huang H, Liaw S. An analysis of learners' intentions toward virtual reality learning based on constructivist and technology acceptance approaches. IRRODL 2018 Feb 23;19(1):91-115. [doi: 10.19173/irrodl.v19i1.2503]
47. Mäkinen H, Haavisto E, Havola S, Koivisto J. User experiences of virtual reality technologies for healthcare in learning: an integrative review. Behav Inform Technol 2020 Jul 01;41(1):1-17. [doi: 10.1080/0144929x.2020.1788162]
48. Bond CF, Titus LJ. Social facilitation: a meta-analysis of 241 studies. Psychol Bull 1983;94(2):265-292 [FREE Full text] [doi: 10.1037/0033-2909.94.2.265] [Medline: 32612559]
49. Zajonc RB. Social facilitation. Science 1965 Jul 16;149(3681):269-274. [doi: 10.1126/science.149.3681.269]
50. Strojny PM, Dużmańska-Misiarczyk N, Lipp N, Strojny A. Moderators of social facilitation effect in virtual reality: co-presence and realism of virtual agents. Front Psychol 2020;11:1252 [FREE Full text] [doi: 10.3389/fpsyg.2020.01252] [Medline: 32612559]
51. Meri-Yilan S. A Constructivist Desktop Virtual Reality-Based Approach to Learning in a Higher Education: Emerging Technologies in Virtual Learning Environments Institution. Hershey, PA: IGI Global; 2019:258-283.
52. Breitkreuz KR, Kardong-Edgren S, Gilbert GE, DeBlieck C, Maske M, Hallock C, et al. A multi-site study examining the usability of a virtual reality game designed to improve retention of sterile catheterization skills in nursing students. Simul Gaming 2020 Sep 25;52(2):169-184. [doi: 10.1177/1046878120954891]

Abbreviations

ISBAR: introduction, situation, background, assessment, and recommendation
SUS: System Usability Scale
VR: virtual reality
4C/ID: four-component instructional design

Edited by A Kushniruk; submitted 22.06.22; peer-reviewed by J Bagby, J Brooke; comments to author 06.10.22; revised version received 27.10.22; accepted 08.11.22; published 29.12.22.

Please cite as:
Andreasen EM, Høigaard R, Berg H, Steinsbekk A, Haraldstad K
Usability Evaluation of the Preoperative ISBAR (Identification, Situation, Background, Assessment, and Recommendation) Desktop Virtual Reality Application: Qualitative Observational Study
JMIR Hum Factors 2022;9(4):e40400
URL: https://humanfactors.jmir.org/2022/4/e40400
doi:10.2196/40400
PMID:

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