Original Paper

Effects of Video-Based Patient Education and Consultation on Unplanned Health Care Utilization and Early Recovery After Coronary Artery Bypass Surgery (IMPROV-ED): Randomized Controlled Trial

Gijs van Steenbergen[1], MD; Dennis van Veghel[1], PhD; Dideke van Lieshout[2], MA; Merel Sperwer[2], MSc; Joost ter Woorst[1], MD, PhD; Lukas Dekker[1,3], MD, PhD

[1]Cardiothoracic Surgery Department, Catharina Heart Centre, Catharina Hospital, Eindhoven, Netherlands

[2]Dutch Heart Foundation, The Hague, Netherlands

[3]Department of Biomedical Technology, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Gijs van Steenbergen, MD
Cardiothoracic Surgery Department
Catharina Heart Centre
Catharina Hospital
Michaelangelolaan 3
PO Box 1350
Eindhoven, 5602
Netherlands
Phone: 31 040 239 9111
Email: gijs.v.steenbergen@catharinaziekenhuis.nl

Abstract

Background: Health care utilization after coronary artery bypass graft (CABG) surgery is high and is partly of an unplanned nature. eHealth applications have been proposed to reduce care consumption, which involve and assist patients in their recovery. In this way, health care expenses could be reduced and quality of care could be improved.

Objective: The aim of this study was to evaluate if an eHealth program can reduce unplanned health care utilization and improve mental and physical health in the first 6 weeks after CABG surgery.

Methods: A single-blind randomized controlled trial was performed, in which patients scheduled for nonacute CABG surgery were included from a single center in the Netherlands between February 2020 and October 2021. Participants in the intervention group had, alongside standard care, access to an eHealth program consisting of online education videos and video consultations developed in conjunction with the Dutch Heart Foundation. The control group received standard care. The primary outcome was the volume and costs of a composite of unplanned health care utilization, including emergency department visits, outpatient clinic visits, rehospitalization, patient-initiated telephone consultations, and visits to a general practitioner, measured using the Medical Technology Assessment Medical Consumption Questionnaire. Patient-reported anxiety and recovery were also assessed. Intention-to-treat and “users-only” analyses were used.

Results: During the study period, 280 patients were enrolled and randomly allocated at a 1:1 ratio to the intervention or control group. The intention-to-treat analysis consisted of 136 and 135 patients in the intervention and control group, respectively. At 6 weeks, the primary endpoint had occurred in 43 of 136 (31.6%) patients in the intervention group and in 61 of 135 (45.2%) patients in the control group (hazard ratio 0.56, 95% CI 0.34-0.92). Recovery was faster in the intervention group, whereas anxiety was similar between study groups. “Users-only” analysis yielded similar results.

Conclusions: An eHealth strategy comprising educational videos and video consultations can reduce unplanned health care utilization and can aid in faster patient-reported recovery in patients following CABG surgery.

Trial Registration: Netherlands Trial Registry NL8510; https://trialsearch.who.int/Trial2.aspx?TrialID=NL8510

International Registered Report Identifier (IRRID): RR2-10.1007/s12471-020-01508-9

(J Med Internet Res 2022;24(8):e37728) doi: 10.2196/37728

KEYWORDS

e-Health; eHealth; digital health; patient education; coronary artery bypass surgery; cardiac surgery; health care utilization; costs; cost; economic; coronary; cardiology; heart; surgery; bypass; RCT; randomized controlled trial; video consultation; telehealth; telemedicine; patient-reported; recovery; expense

Introduction

Coronary artery bypass graft (CABG) surgery is one of the most frequently performed cardiac surgeries in the world, which is generally performed with good outcomes and relatively low 30-day mortality (~1.5%) [1]. In more recent years, the care chain for patients undergoing CABG surgery has been demonstrated to increase efficiency and reduce costs. As a result, the duration of hospitalization has decreased substantially, with patients discharged on the 7th postoperative day (mean). These efficiency-driven early discharge protocols require more self-management skills among patients. Early discharge reduces the time physicians can spend with their patients in the direct postoperative phase in spite of the well-known benefit of patient counseling and guidance through recovery [2,3].

After discharge, patients commonly experience anxiety or uncertainty about symptoms or appropriate physical exercise [4]. These issues are typically addressed during hospitalization; however, after discharge, patients' recall of information is often incomplete and they do not always know who to address with questions [4]. The advantages of a shortened hospital stay might therefore be counterbalanced by preventable unplanned health care utilization, especially since planned care is not initiated until several weeks after discharge. At present, nearly 1 in 7 patients are readmitted in the first 30 days after discharge for noncardiac causes and roughly 15% of patients visit the emergency department within 1 month after CABG surgery [5-8]. It was estimated that potentially preventable readmissions following CABG surgery cost Medicare US $151 million in 2005, placing a significant burden on society [7]. With the expected increase in the number of future patients undergoing CABG surgery, this is a pressing issue urging evaluation and a potential redesign of postoperative follow-up.

eHealth is defined by the World Health Organization as "the cost-effective and secure use of information and communication technologies in support of health and health-related fields," which encompasses multiple digital interventions that can aid in the delivery of patient-centered care and postoperative patient guidance, thereby potentially reducing unplanned health care utilization [9]. eHealth strategies have been successfully applied in postoperative follow-up in various forms, which have been shown to improve patient outcomes, speed recovery, and reduce health care utilization in various surgical populations [10]. In addition, eHealth has proven to be of value for patients to enhance their self-management through better understanding of their disease, increased independence, and improved acceptance to adhere to lifestyle advice [3,11]. However, experience with eHealth in patients following CABG surgery is limited, and it remains unclear if eHealth strategies would be effective in this population.

The objective of this trial was to fill this knowledge and experience gap. We hypothesized that restructuring the postoperative period with an eHealth strategy will reduce unplanned health care utilization through improved mental and physical health and faster recovery.

Methods

Trial Design

The IMPROV-ED trial was a randomized controlled trial (RCT) performed between February 2020 and December 2021 at Catharina Hospital in the Netherlands. A detailed study protocol was published prior to enrollment of the first study participant [11]. No changes were made to the study protocol between publication and initiation of the trial. The trial is reported using the CONSORT (Consolidated Standards of Reporting Trials) checklist for RCTs [12].

Ethics Considerations

The study was approved by the medical ethics committee (R19.100) and was registered in the Netherlands Trial Registry (NL8510). Written informed consent was obtained from all patients who met the inclusion criteria and were willing to participate.

Participants

To minimize selection bias, all patients on the waiting list for isolated CABG surgery over 18 years of age were contacted by telephone and informed about the study by one of the investigators. Patients were eligible for participation if they had access to a computer/tablet/smartphone with internet connection and a webcam/built-in camera; had sufficient knowledge of the use of internet and email (assistance was allowed); and were able to speak, read, and interpret the Dutch language. The eHealth strategy would not be applicable to patients who did not comply with these inclusion criteria and they were therefore not eligible for participation. At inclusion, patients were randomized 1:1 to the intervention or control group using a block size of 4. A certified program was used for sequence generation and randomization (Research Manager). When a patient was randomized but no longer qualified for the inclusion criteria or was lost to follow-up, the patient was excluded from further follow-up and analysis.

Interventions

Patients randomized to the control group received standard postoperative care, comprising planned outpatient follow-up by their cardiothoracic surgeon at 6-8 weeks after discharge and a cardiac rehabilitation program supervised by cardiologists with outpatient follow-up starting between 4 and 8 weeks after surgery. As a result of the COVID-19 health crisis and the measures taken by the government, most of these contacts were telephone consultations (TCs) rather than physical consultations.

Patients randomized to the intervention group had access to the eHealth strategy in addition to standard care. The eHealth strategy comprised web-based educational videos developed by the Dutch Heart Foundation and two postoperative video consultations (VCs) with a physician from the department of cardiothoracic surgery at 1 and 3 weeks after discharge.

Upon randomization, patients in the intervention group received access to the educational videos via a link sent by email. The same link was sent via email again at discharge. By clicking the link, patients were referred to a hidden (for nonparticipants and the control group) part of the website from the Dutch Heart Foundation that contained the educational videos. The content of the educational videos was constructed and validated by physicians and patient representatives prior to the trial. Based on these evaluations and a scoping review of the literature on delivery of information to patients with varying degrees of health literacy, the full content was delivered to patients at inclusion instead of by fragmentized access to videos applicable to the patient's situation [13]. Nevertheless, to prevent cognitive overload in patients with low health literacy, educational videos were categorized in three categories: treatment (10 videos with information on the surgery and how to prepare for admission), recovery (6 videos about what to expect in the postoperative course and when to contact a physician), and healthy living (2 videos on cardiovascular risk management, including smoking cessation, weight reduction, cholesterol management, and exercise). The videos were delivered in spoken text supported by animations for optimal health communication to patients with low and adequate health literacy [13]. Usage data were extracted from the web log for evaluation purposes. Educational videos were available to patients in the intervention group throughout the trial (ie, not only when the link was sent). See the published study protocol for an illustrative overview of the educational videos [11].

VCs were conducted by a nurse practitioner or junior doctor using Microsoft Teams. The dates for VCs were sent to patients by email at discharge. On the day of the VC, patients received an email with a link providing access to the VC. The VC was not scheduled on the same day as routine outpatient follow-up. During the VCs, patients were questioned about their recovery and physical and mental complaints. The sternotomy wound was visually inspected. Patients who required physical examination or diagnostic tests based on the VC were instructed to visit the general practitioner or emergency department, or were scheduled for early outpatient follow-up (within 1 week) at discretion of the physician. The nurse practitioner/junior doctor who conducted the VCs was blinded to the study's objectives and outcomes. Study participants were not blinded. If a VC was unexpectedly not possible (eg, due to unforeseen connection errors, problems with hardware, technical issues), the VC was replaced by a TC. Reasons for replaced VC were reported.

Outcomes

The primary outcomes of the IMPROV-ED trial were the volume and costs of unplanned health care utilization defined by a composite of all emergency department visits, outpatient clinic visits, rehospitalization, patient-initiated TCs, and visits to a general practitioner, as measured by the Institute for Medical Technology Assessment Medical Consumption Questionnaire (iMCQ) at the 6-week follow-up [14]. Cross-validation with the patients' reported health care utilization was performed by contacting their health care providers. The secondary outcomes were the individual unplanned health care activities, and a composite of planned and unplanned in-hospital care (emergency department visits, outpatient clinic visits, rehospitalization, and patient-initiated TCs) and planned and unplanned primary care (consultations with a general practitioner, allied health professionals, psychologists) at 6 weeks. The other secondary outcomes were the patients' self-reported physical and mental health, as measured with the Hospital Anxiety and Depression Scale (HADS) and Recovery Index-10 (RI-10) questionnaires [15,16].

Data Collection

All patients received questionnaires at inclusion (anxiety subscale of the HADS), at discharge (HADS and RI-10), 1 week after discharge (HADS and RI-10), 2 weeks after discharge (HADS and RI-10), and 6 weeks after discharge (HADS, RI-10, and iMCQ). Only the anxiety subscale from the HADS was used. A higher score indicated more symptoms of anxiety (HADS maximum score 21) or favorable progress of recovery (RI-10 maximum score 50). The iMCQ resulted in absolute frequencies of visits for the questioned care activities. Patients in the intervention group also received a self-made questionnaire to evaluate the eHealth strategy and to question them about the use of the education videos (see Figures S1 and S2 in Multimedia Appendix 1). If patients had not returned the iMCQ by 8 weeks postdischarge, the questionnaire was conducted over the telephone. If patients had not returned 2 subsequent questionnaires, a research nurse called patients with a reminder. Questionnaires that were not returned or collected otherwise were considered missing.

Statistical Analysis

We calculated the sample size needed for the study based on the expected effect of the intervention on the primary outcome. Previous studies using a comparable eHealth strategy in CABG patients with health care utilization measured with the iMCQ were not available. In a study with abdominal surgery patients, total health care utilization was estimated at a mean of 0.88 (SD 0.15) per patient [17]. In a systematic review by van der Meij et al [10], the effect of an eHealth strategy in surgical patients was not consistent. Therefore, a small to medium effect (d=0.35) was expected from our intervention. Combined with an α of .05 and a power of 0.80, a total sample size of 260 patients was required. We aimed for 280 participants to account for loss to follow-up and nonadherence to the intervention and return of questionnaires (attrition rate 5%, rounded up to a whole number). Demographic data of randomized patients were collected using definitions in line with the Netherlands Heart Registration [18]. Education was grouped into three levels (low, medium, and high) according to the general definition by Statistics Netherlands (see Multimedia Appendix 1 for the full definition).

The main analysis was performed according to the intention-to-treat (ITT) principle. Because patients in the intervention group were not obliged to use the educational videos and VCs might not be possible due to technical errors, per-protocol analysis was also performed, which included only patients who used the intervention strategy as intended (defined as having at least one VC or TC and accessed the educational videos at least once).

Planned subgroup analyses of the primary outcome were performed according to age (<65 years vs ≥65 years), sex, recent myocardial infarction, left ventricular function, diabetes, type of CABG (on-pump vs off-pump), log EuroScore, and highest level of education.

Continuous variables and outcomes are expressed as mean (SD) in cases of a normal distribution and as median (IQR) in cases of a nonnormal distribution. The Kolmogorov-Smirnov test and Q-Q plots were used to test for normality of the data distribution. Categorical data are summarized as absolute and relative frequencies. The updated Dutch Manual for Cost Analysis in Health Care Research was used as the source for cost prices per health care activity if available [19]. Other tariffs were calculated using top-down microcosting as described by Tan and Hakkaart-van Roijen et al [20, 21] (see Multimedia Appendix 1 for details). Each consumed health care activity was multiplied by the cost price and total costs were calculated by summing these multiplications. The HADS and RI-10 questionnaire scores at each interval were compared between study groups. *P*<.05 was considered significant. Primary and secondary outcomes are presented with effect-size estimates and 95% CIs using the Cox proportional hazards model. The proportional hazard assumption was assessed by log (–log) plots. Analyses were performed using SPSS 25 and RStudio.

Results

Study Population

In total, 280 patients were included in the study between February 2020 and December 2021, and subsequently randomized yielding 140 patients in each study group. One patient in the intervention group and two patients in the control group were excluded after randomization because they underwent percutaneous coronary intervention instead of CABG surgery. In the intervention group, three patients were lost to follow-up (1 withdrew consent, 1 had an early readmission due to a complication, and 1 died). In the control group, three patients were lost to follow-up (1 withdrew consent and 2 died). The ITT analysis therefore consisted of 136 and 135 patients in the intervention and control group, respectively. Weblog and planning data revealed that 8 patients did not use the intervention as intended, whereby 128 patients were included in the intervention group in the per-protocol analysis (Figure 1).

Baseline characteristics of patients were similar in the two groups (Table 1), with a median age of 67.9 and 69.6 years for the intervention and control group, respectively. The majority of patients were male in both groups. At the time of surgery, 25% of patients had an urgent indication and the remainder underwent surgery in the elective setting. In the majority of patients, on-pump CABG was performed using 3 distal anastomoses. The left or right internal mammary artery was used in >98% of patients. Duration of admission was also similar in the two groups (Table 1).

Figure 1. Enrollment overview. CABG: coronary artery bypass graft; PCI: percutaneous coronary intervention.

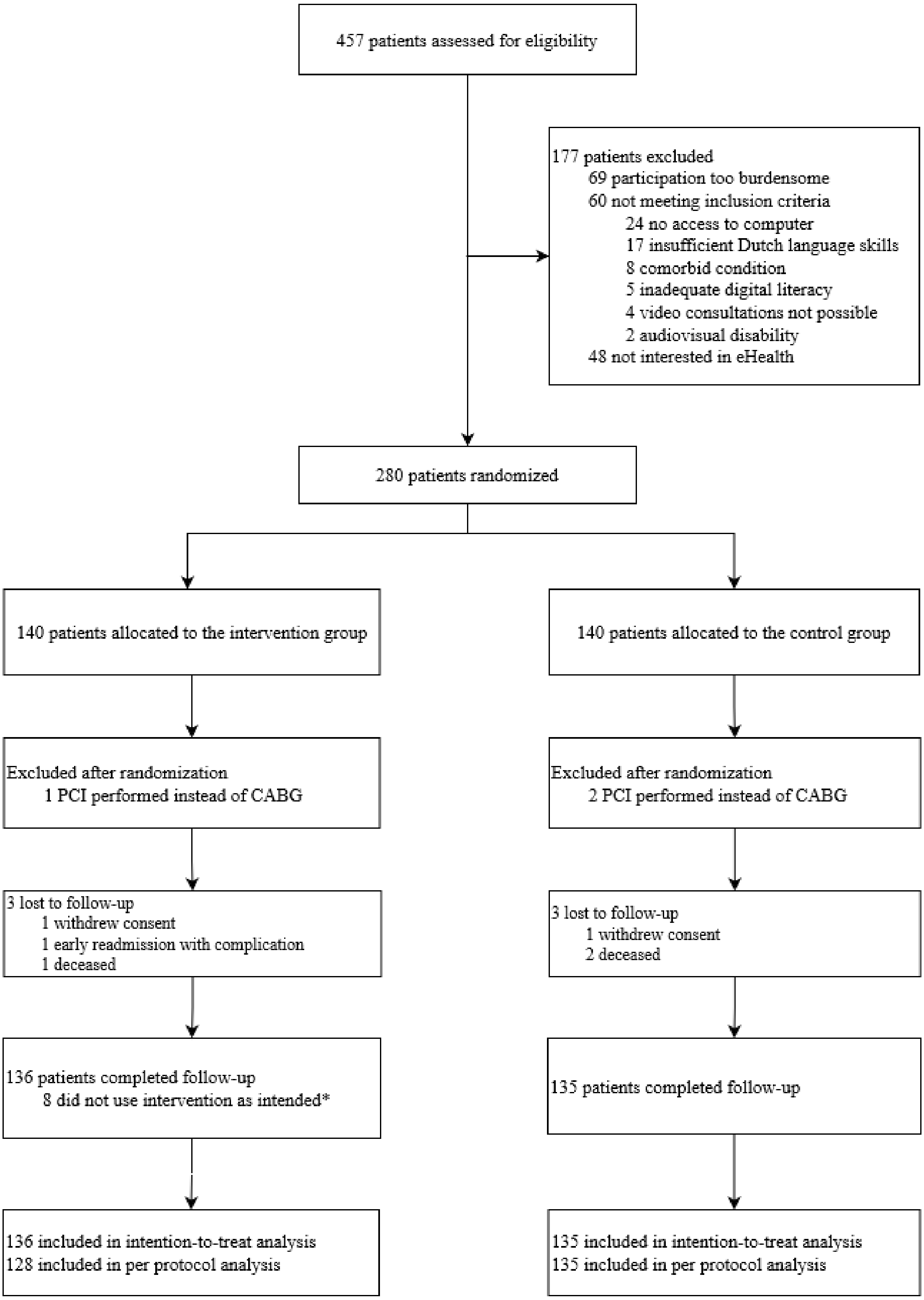


Table 1. Baseline characteristics and procedural data (intention-to-treat analysis).

Characteristics	eHealth group (n=136)	Standard care (n=135)
Age (years), median (IQR)	67.9 (61.5-73.3)	69.6 (65.2-74.1)
Male, n (%)	121 (89.6)	113 (83.1)
BMI, median (IQR)	27.7 (25.1-30.6)	27.2 (25.2-30.3)
Medical history, n (%)		
Diabetes mellitus	45 (33.3)	33 (24.3)
Chronic pulmonary disease	7 (5.2)	15 (11.0)
Atrial fibrillation	9 (6.7)	6 (4.4)
Multivessel disease	117 (86.7)	121 (89.0)
Peripheral vascular disease	17 (12.5)	17 (12.6)
Renal impairment (MDRD[a]<60 mL/min/1.73 m^2)	10 (7.4)	1 (8.1)
Previous stroke	3 (2.2)	5 (3.7)
Recent MI[b] (90 days)	45 (33.3)	46 (33.8)
Previous PCI[c]	36 (26.5)	31 (22.9)
Left ventricular ejection fraction, median (IQR)	55 (50-55)	55 (50-55)
Ejection fraction≤30%, n (%)	0 (0)	3 (2.2)
NYHA[d] class>II, n (%)	3 (2.2)	7 (5.2)
Current health status		
SF-36[e] physical score, median (IQR)	51 (43-56)	48 (40-51)
SF-36 mental score, median (IQR)	58 (55-63)	59 (55-64)
HADS[f], median (IQR)	3 (1-7)	3 (1-6)
Level of education, n (%)		
Low	36 (26.5)	42 (31.1)
Intermediate	53 (39.0)	55 (40.7)
High	47 (34.6)	38 (28.1)
Procedural data		
EuroSCORE log, median (IQR)	2.40 (1.82-4.06)	2.87 (2.01-4.28)
EuroSCORE II, median (IQR)	1.41 (1.05-2.04)	1.32 (0.78-2.43)
Use of ECC[g], n (%)	110 (81.5)	101 (74.8)
ECC duration in users (min), median (IQR)	74 (60-91)	76 (64-91)
Number of distal anastomoses, median (IQR)	3 (2-4)	3 (2-4)
Hospital stay (days), median (IQR)	6 (5-7)	6 (5-7)

[a] MDRD: Modification of Diet in Renal Disease.

[b] MI: myocardial infarction.

[c] PCI: percutaneous coronary intervention.

[d] NYHA: New York Heart Association.

[e] SF-36: Short Item-36.

[f] HADS: Hospital Anxiety and Depression Scale.

[g] ECC: extracorporeal circulation.

Outcomes

At 6 weeks, care was consumed by less patients in the intervention group than in the control group (Table 2). The benefit of the eHealth strategy was most noticeable in patients over 65 years of age, those of male sex, those with recent myocardial infarction, or with a EuroScore>2 (see Figure S3 in Multimedia Appendix 1). Reduction in individual care activities was significantly different between groups for TCs and was borderline significant for general practitioner visits (Table 2). Costs related to the primary outcome were significantly higher in the standard care group compared with those in the eHealth group (*P*<.001, Table 2), which was attributed to the higher volume of care consumption in the control group (see Table S1 in Multimedia Appendix 1).

A composite of unplanned in-hospital care, a composite of planned and unplanned in-hospital care after discharge, and use of planned and unplanned primary care were all higher in the control group than the intervention group (Table 2). The volume of consumed care was also higher in the control group (Table S1 of Multimedia Appendix 1).

The RI-10 score, indicating patient-reported recovery, was significantly higher in the intervention group in the 3rd and 6th weeks after discharge (Figure 2). Anxiety was not significantly different between study groups (Figure 2). Per-protocol analysis revealed similar findings (see Tables S2 and S3 in Multimedia Appendix 1).

Table 2. Outcomes at 6 weeks.

Outcomes	eHealth group (n=136)	Standard care (n=135)	Hazard ratio (95% CI)	*P* value
Primary outcomes				
Composite outcome[a], n (%)	43 (31.6)	61 (45.2)	0.56 (0.34-0.92)	.02
Cost (Euro[b]), Median (IQR)	0 (0-95)	66 (0-215)	N/A[c]	<.001
Cost (Euro), mean (SD)	183 (515)	285 (777)	N/A	<.001
Secondary outcomes, n (%)				
Composite unplanned in-hospital care	36 (26.5)	53 (39.3)	0.56 (0.33-0.93)	.03
Emergency department visits	14 (10.3)	23 (17.0)	0.56 (0.27-1.14)	.11
Readmissions	7 (5.1)	9 (6.7)	0.76 (0.28-2.10)	.59
Outpatient clinic visits	11 (8.1)	10 (7.4)	1.10 (0.45-2.68)	.83
Telephone consultations	29 (21.3)	47 (34.8)	0.51 (0.29-0.87)	.01
General practitioner visits (unplanned)	28 (20.6)	41 (30.4)	0.59 (0.34-1.04)	.07
Composite of all in-hospital care[d]	69 (50.7)	97 (71.9)	0.40 (0.24-0.67)	<.001
Composite of all primary care[e]	82 (60.3)	101 (74.8)	0.58 (0.36-0.97)	.04

[a] Composite of unplanned health care utilization (ie, emergency department visits, readmissions, outpatient clinic visits, telephone consultations, or general practitioner visits).

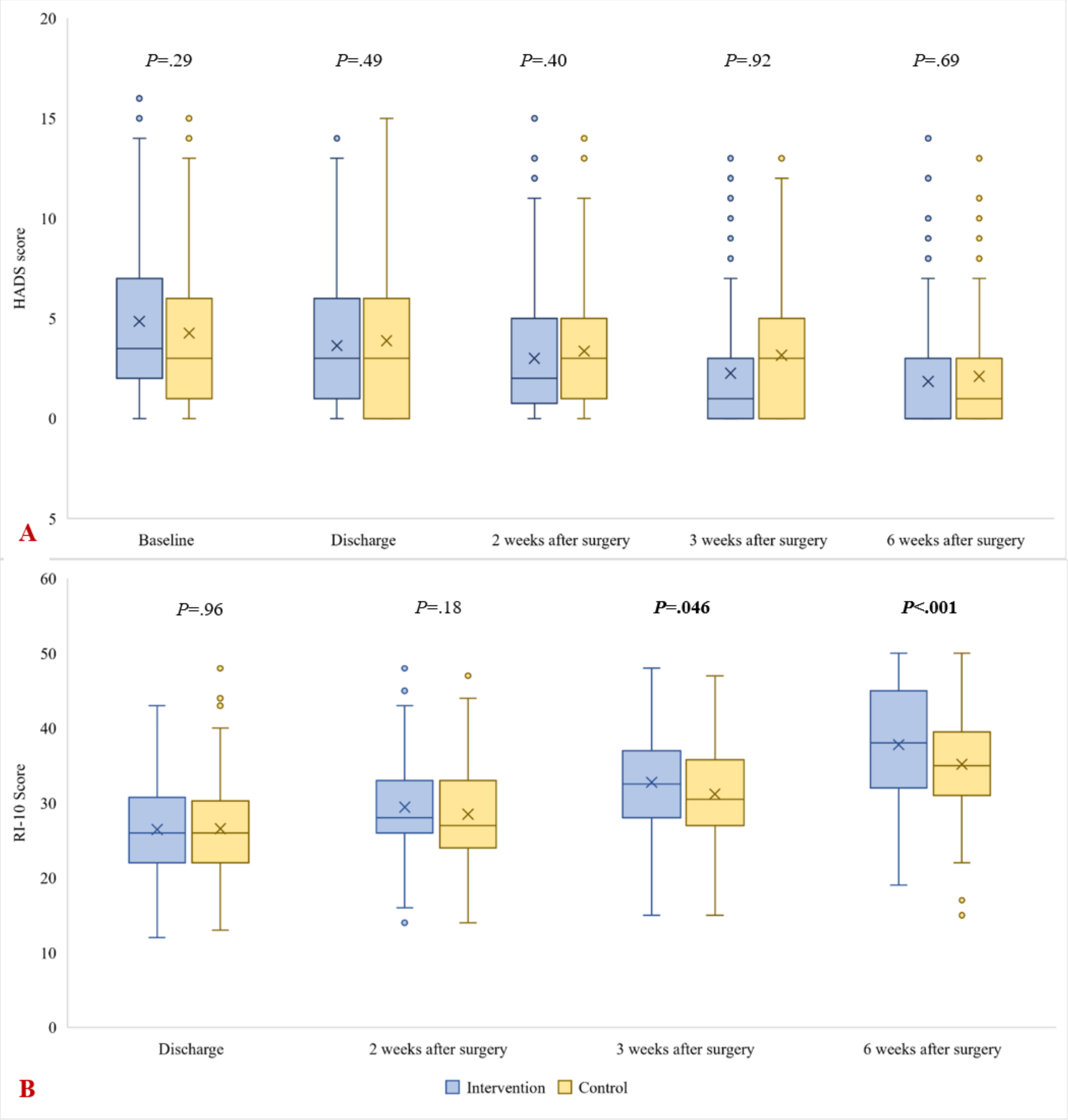
[b] 1 Euro=US $1.13.

[c] N/A: not applicable.

[d] Composite of in-hospital care comprising planned and unplanned emergency department visits, readmissions, outpatient clinic visits, and telephone consultations.

[e] Composite of primary care comprising planned and unplanned visits to the general practitioner, visits to allied health professionals (physical therapists, dieticians, speech therapists, exercise therapy, social workers), and psychologist visits.

Figure 2. Anxiety level measured with the Hospital Anxiety and Depression Scale (HADS) anxiety subscale (A) and progress of recovery measured using the recovery index-10 (RI-10) questionnaire (B).



Process Evaluation

Among patients who were provided access to the educational videos, 95% accessed the videos at least once. A total of 248 of the 272 planned VCs were conducted. Eight patients did not use the intervention as intended and did not receive VCs (n=16 VCs). The other VCs that did not take place were substituted with a TC due to technical errors (n=8). The median duration of VCs was 10 minutes (IQR 8-11) for the first VC and was 8 minutes (IQR 7-9) for the second VC. Patients reported positive attitudes toward the education videos and the VC (Figures S1 and S2 in Multimedia Appendix 1). Most notably, patients reported a fairly positive attitude toward substitution of a face-to-face contact with a VC, but patients also reported that the VC with a physician (not the surgeon) or nurse practitioner should not substitute the consultation with the surgeon.

Discussion

Principal Findings

The principal finding of the IMPROV-ED trial is that an eHealth strategy comprising educational videos and VCs results in a reduction of unplanned care and costs. In addition, the eHealth strategy is associated with faster patient-reported recovery. These findings are of medical and societal importance given the increasing interest in digital health and the need for value-based alongside evidence-based care. Our study is the first to provide robust evidence that an eHealth intervention can aid in reduction of health care utilization and associated costs. This effect appears applicable to both in-hospital care as well as primary care. One of the most pressing concerns from health care insurance companies and decision-makers toward eHealth is the great investment that is required for development of content and the necessary infrastructure and issues that arise after implementation due to lack of reimbursement options [22]. Our findings refute these concerns by showing positive effects on costs. Furthermore, the eHealth strategy did not only contribute to less patients consuming care (Table 2) but also reduced the care consumed per patient (Table S1 of Multimedia Appendix 1), which underlines the high potential of eHealth strategies for this patient population to also positively influence the burden on health care personnel. With an aging population, a vast increase in health care consumption is expected in the near future. Based on the results of our study, an eHealth program is proven to aid in the sustainment of health care systems.

The findings of our study shine new light on previous studies comparing an eHealth program with standard care because it is the first study to use health care utilization as a primary outcome [10]. Very limited studies are available that use health care utilization as an outcome, and those that have considered care consumption as a secondary or tertiary outcome. Previous studies were also not adequately powered to draw reliable conclusions on the use of eHealth in reduction of care consumption and, consequently, these studies reported mixed outcomes. For example, Keeping-Burke et al [23] incorporated health care use as a tertiary outcome in an RCT of patients after CABG surgery using postoperative VCs, and concluded that patients in the telehealth group had fewer physician contacts. Zahlmann et al [24] used telecommunication in the postoperative period after cataract surgery (n=62) and also concluded that care in the intervention group was lower than that in the control group. Conversely, Barnason et al [25,26] conducted two RCTs in 232 and 50 CABG patients, respectively, using a supportive telehealth program and concluded that both groups had similar health care use at 6-month follow-up. Barnason et al [25] and Keeping-Burke et al [23] both reported no differences in emergency department visits and readmissions between study groups. Readmission was also similar in a study by Gandsas et al [27] after laparoscopic gastric bypass using robotic telerounding during admission.

Another major strength of the current eHealth program is that it provides patients of various degrees of socioeconomic status and health literacy with information on the procedure and their medical condition from a reliable source that is endorsed by their surgeon. The Dutch Heart Foundation is a respected organization that is dedicated to providing information on cardiovascular health, advocating patient interests, and conducting research [28]. The educational videos are developed in conjunction with patient representatives and physicians. In the VCs, additional questions are answered and uncertainties are addressed. The impact of educational videos and VCs is presumably in improvement of self-management skills and reduction of fear and anxiety. Recall of information on information provided preoperatively or at discharge is often incomplete, and patients might not know what physical activity is allowed after discharge or who to contact in case of complaints. Patients can turn to the internet for information; however, this information is uncontrolled, sometimes inaccurate, and is not tailored to the care processes of their provider. Because planned care is not initiated until 6 weeks after surgery (and sometimes later in practice), conflicting advice can induce insecurity, which will lead to use of care and will hamper recovery. The results of our study are consistent with this hypothesis. Nevertheless, the anxiety symptoms measured with the HADS questionnaire relate to anxiety in a narrow sense, whereas the anxiety experienced by patients after CABG surgery is likely to be more subtle in nature, which may have contributed to the nonsignificant difference in measured anxiety found in this study.

However, health care utilization is the resultant of a multifactorial behavioral model that attributes a combination of predisposing factors (eg, patient characteristics such as age, sex, sociodemographic parameters, or health literacy and attitude toward health), enabling factors (eg, income, health insurance status, health care organization), and need factors (eg, experience with health care) to health care utilization [29]. The eHealth strategy used in the IMPROV-ED trial has a positive influence on some of these attributes but not all. Interestingly, subgroup analysis showed that the eHealth program had a greater benefit in more vulnerable patients (EuroScore≥2) and revealed a trend toward more benefit in patients with a low level of education. By contrast, a small group of patients who provided informed consent did not use the educational videos or VCs that were part of the eHealth strategy. These patients reported to have received sufficient information from their physician, nurse, or paramedic during admission, or that they found the relevant information online themselves. It might therefore be reasonable to consider adding different modes of digital health delivery to the currently used eHealth strategy (eg, mobile apps, live chat, home monitoring, telerehabilitation) to manage more attributes of health utilization and to offer a more individualized approach tailored to the patients' needs. Combining different modes of digital care might thereby further reduce health care utilization and potentially also improve clinical outcomes [22].

Learning Points and Limitations

Even though the IMPROV-ED study yielded positive results toward the primary outcome (Table 2) and patients were generally positive about the eHealth strategy (Figures S2 and S3 in Multimedia Appendix 1), several learning points and limitations should be taken into account for future eHealth programs.

First, the IMPROVE-ED trial is designed for patients who consume care as a result of insecurity, anxiety, lack of medical knowledge, and/or inadequate discharge counseling. As can be concluded from Figure 1, a relevant number of patients who were invited to participate in the trial did not provide informed consent due to the general burden of having to undergo cardiac surgery (patients used terms such as "stressful," "anxiety," and "insecurity") in conjunction with study obligations. The effect of the eHealth strategy may be underestimated because this group of patients might have been part of the target population in which the eHealth strategy would have incremental value.

Due to ethical constraints (patients did not provide informed consent for participation and thus for data collection), these patients were not further analyzed for the study outcomes.

In this study, standard care was not replaced by digital alternatives, and yet the costs of the intervention group were still lower than those of the control group receiving only standard care. Because VCs were used as an add-on to standard care, there are potentially more opportunities to reduce costs further. The fact that eHealth is being implemented *on top* of current health care services is, in addition to cost concerns, one of the challenges identified by the European Society of Cardiology as hampering the introduction of eHealth into everyday clinical practice [22]. Future endeavors should focus on investigating the potential of substitution of standard physical care with digital alternatives, especially since the patients' attitude was generally positive toward the (hypothetical) substitution of a physical contact with a VC in this study (Figure S1 of Multimedia Appendix 1). Previous studies also stated that it is feasible to obtain the same effective communication and interaction with VCs as with face-to-face care [30].

The majority of patients included in the IMPROV-ED trial were included during the COVID-19 pandemic. The results of the study might therefore be an underrepresentation of care consumption because patients feared transmission in the hospital setting [31]. Nevertheless, the randomized design balances this influence between the study groups.

Conclusion

An eHealth strategy comprising educational videos and VCs can reduce unplanned in-hospital and primary health care utilization and costs, and can aid in faster patient-reported recovery following CABG surgery.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Definitions, tariffs for consumed care, and supplementary data (Tables S1-S3, Figures S1-S3).
[DOCX File , 307 KB-Multimedia Appendix 1]

Multimedia Appendix 2

CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 116 KB-Multimedia Appendix 2]

References

1. Weisse AB. Cardiac surgery: a century of progress. Tex Heart Inst J 2011;38(5):486-490 [FREE Full text] [Medline: 22163121]
2. Vonk Noordegraaf A, Anema JR, Louwerse MD, Heymans MW, van Mechelen W, Brölmann HAM, et al. Prediction of time to return to work after gynaecological surgery: a prospective cohort study in the Netherlands. BJOG 2014 Mar;121(4):487-497. [doi: 10.1111/1471-0528.12494] [Medline: 24245993]
3. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
4. Lie I, Bunch EH, Smeby NA, Arnesen H, Hamilton G. Patients' experiences with symptoms and needs in the early rehabilitation phase after coronary artery bypass grafting. Eur J Cardiovasc Nurs 2012 Mar;11(1):14-24. [doi: 10.1016/j.ejcnurse.2010.09.004] [Medline: 21030311]
5. Fox JP, Suter LG, Wang K, Wang Y, Krumholz HM, Ross JS. Hospital-based, acute care use among patients within 30 days of discharge after coronary artery bypass surgery. Ann Thorac Surg 2013 Jul;96(1):96-104 [FREE Full text] [doi: 10.1016/j.athoracsur.2013.03.091] [Medline: 23702228]
6. Shawon MSR, Odutola M, Falster MO, Jorm LR. Patient and hospital factors associated with 30-day readmissions after coronary artery bypass graft (CABG) surgery: a systematic review and meta-analysis. J Cardiothorac Surg 2021 Jun 10;16(1):172 [FREE Full text] [doi: 10.1186/s13019-021-01556-1] [Medline: 34112216]
7. Report to the Congress: promoting greater efficiency in Medicare. Medicare Payment Advisory Commission. 2007. URL: https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/Jun07_ExecSummary.pdf [accessed 2021-10-03]
8. Saab S, Noureddine S, Dumit NY. Readmission rates and emergency department visits after coronary artery bypass graft surgery and related factors. J Med Liban 2013;61(3):155-160. [doi: 10.12816/0001444] [Medline: 24422366]
9. Fifty-eighth World Health Assembly, Geneva, 16-25 May 2005: resolutions and decisions: annex. World Health Organization. 2005. URL: https://apps.who.int/iris/handle/10665/20398 [accessed 2020-02-01]
10. van der Meij E, Anema JR, Otten RHJ, Huirne JAF, Schaafsma FG. The effect of perioperative e-health interventions on the postoperative course: a systematic review of randomised and non-randomised controlled trials. PLoS One 2016;11(7):e0158612 [FREE Full text] [doi: 10.1371/journal.pone.0158612] [Medline: 27383239]

11. van Steenbergen GJ, van Veghel D, Ter Woorst J, van Lieshout D, Dekker L. IMPROV-ED trial: eHealth programme for faster recovery and reduced healthcare utilisation after CABG. Neth Heart J 2021 Feb;29(2):80-87 [FREE Full text] [doi: 10.1007/s12471-020-01508-9] [Medline: 33141398]
12. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
13. Meppelink C. Designing digital health information in a health literacy context. PhD Thesis. University of Amsterdam. 2016. URL: https://dare.uva.nl/search?identifier=220fb9ef-1a51-4325-9019-61cfdaf8633f [accessed 2022-07-21]
14. Manual iMTA Medical Cost Questionnaire (iMCQ). iMTA Productivity Health Research Group. URL: https://www.imta.nl/questionnaires/imcq/ [accessed 2022-07-21]
15. Kluivers KB, Hendriks JCM, Mol BWJ, Bongers MY, Vierhout ME, Brölmann HAM, et al. Clinimetric properties of 3 instruments measuring postoperative recovery in a gynecologic surgical population. Surgery 2008 Jul;144(1):12-21. [doi: 10.1016/j.surg.2008.03.027] [Medline: 18571580]
16. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med 1997 Mar;27(2):363-370. [doi: 10.1017/s0033291796004382] [Medline: 9089829]
17. Pattamatta M, Smeets BJJ, Evers SMAA, Peters EG, Luyer MDP, Hiligsmann M. Quality of life and costs of patients prior to colorectal surgery. Expert Rev Pharmacoecon Outcomes Res 2020 Apr;20(2):193-198. [doi: 10.1080/14737167.2019.1628641] [Medline: 31190575]
18. De Nederlandse Hart Registratie (Netherlands Heart Registration). 2020. URL: https://nederlandsehartregistratie.nl/handboeken/ [accessed 2020-02-01]
19. Kanters TA, Bouwmans CAM, van der Linden N, Tan SS, Hakkaart-van Roijen L. Update of the Dutch manual for costing studies in health care. PLoS One 2017;12(11):e0187477 [FREE Full text] [doi: 10.1371/journal.pone.0187477] [Medline: 29121647]
20. Tan SS. Microcosting in economic evaluations: issues of accuracy, feasibility, consistency and generalisability. Erasmus University. 2009. URL: http://hdl.handle.net/1765/17354 [accessed 2020-02-01]
21. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan SS. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg In opdracht van Zorginstituut Nederland Geactualiseerde versie. Kostenhandleiding 2015.
22. Frederix I, Caiani EG, Dendale P, Anker S, Bax J, Böhm A, et al. ESC e-Cardiology Working Group Position Paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
23. Keeping-Burke L, Purden M, Frasure-Smith N, Cossette S, McCarthy F, Amsel R. Bridging the transition from hospital to home: effects of the VITAL telehealth program on recovery for CABG surgery patients and their caregivers. Res Nurs Health 2013 Dec;36(6):540-553. [doi: 10.1002/nur.21571] [Medline: 24242195]
24. Zahlmann G, Mertz M, Fabian E, Holle R, Kaatz H, Neubauer L, et al. Perioperative cataract OP management by means of teleconsultation. Graefes Arch Clin Exp Ophthalmol 2002 Jan;240(1):17-20. [doi: 10.1007/s00417-001-0396-0] [Medline: 11954775]
25. Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M, et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. Heart Lung 2009;38(5):364-376 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.005] [Medline: 19755186]
26. Barnason S, Zimmerman L, Schulz P, Tu C. Influence of an early recovery telehealth intervention on physical activity and functioning after coronary artery bypass surgery among older adults with high disease burden. Heart Lung 2009;38(6):459-468 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.010] [Medline: 19944870]
27. Gandsas A, Parekh M, Bleech MM, Tong DA. Robotic telepresence: profit analysis in reducing length of stay after laparoscopic gastric bypass. J Am Coll Surg 2007 Jul;205(1):72-77. [doi: 10.1016/j.jamcollsurg.2007.01.070] [Medline: 17617335]
28. Nederlandse Hartstichting (Dutch Heart Foundation). URL: https://www.hartstichting.nl/ [accessed 2020-02-06]
29. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998-2011. Psychosoc Med 2012;9:Doc11. [doi: 10.3205/psm000089] [Medline: 23133505]
30. Jiménez-Rodríguez D, Ruiz-Salvador D, Rodríguez Salvador MDM, Pérez-Heredia M, Muñoz Ronda FJ, Arrogante O. Consensus on criteria for good practices in video consultation: a Delphi study. Int J Environ Res Public Health 2020 Jul 27;17(15):5396 [FREE Full text] [doi: 10.3390/ijerph17155396] [Medline: 32727042]
31. Derks L, Sturkenboom HN, Zaal M, Houterman S, Woudstra P, Tio RA, et al. Association between public media and trends in new acute coronary syndrome presentations during the first COVID-19 wave in the Netherlands. Neth Heart J 2021 Nov;29(11):577-583 [FREE Full text] [doi: 10.1007/s12471-021-01603-5] [Medline: 34327671]

Abbreviations

CABG: coronary artery bypass graft

CONSORT: Consolidated Standards of Reporting Trials
HADS: Hospital Anxiety and Depression Scale
iMCQ: Institute for Medical Technology Assessment Medical Consumption Questionnaire
ITT: intention to treat
RCT: randomized controlled trial
RI-10: Recovery Index-10
TC: telephone consultation
VC: video consultation

Edited by G Eysenbach; submitted 04.03.22; peer-reviewed by E van der Velde, K Matthias, M Nomali; comments to author 06.05.22; revised version received 04.07.22; accepted 18.07.22; published 26.08.22

Please cite as:
van Steenbergen G, van Veghel D, van Lieshout D, Sperwer M, ter Woorst J, Dekker L
Effects of Video-Based Patient Education and Consultation on Unplanned Health Care Utilization and Early Recovery After Coronary Artery Bypass Surgery (IMPROV-ED): Randomized Controlled Trial
J Med Internet Res 2022;24(8):e37728
URL: https://www.jmir.org/2022/8/e37728
doi: 10.2196/37728
PMID:

Original Paper

Patients' Willingness to Provide Their Clinical Data for Research Purposes and Acceptance of Different Consent Models: Findings From a Representative Survey of Patients With Cancer

Anja Köngeter[1], MA; Christoph Schickhardt[2], DPhil; Martin Jungkunz[2], DPhil; Susanne Bergbold[3], Dipl; Katja Mehlis[1*], DPhil; Eva C Winkler[1*], DPhil, Prof Dr

[1]Section for Translational Medical Ethics, Department of Medical Oncology, National Center for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany

[2]Section for Translational Medical Ethics, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany

[3]Epidemiological Cancer Registry Baden-Württemberg, German Cancer Research Center, Heidelberg, Germany

[*]these authors contributed equally

Corresponding Author:
Eva C Winkler, DPhil, Prof Dr
Section for Translational Medical Ethics
Department of Medical Oncology
National Center for Tumor Diseases, Heidelberg University Hospital
Im Neuenheimer Feld 280
Heidelberg, 69120
Germany
Phone: 49 6221 5636049
Email: eva.winkler@nct-heidelberg.de

Abstract

Background: Secondary use of clinical data for biomedical research purposes holds great potential for various types of noninterventional, data-driven studies. Patients' willingness to support research with their clinical data is a crucial prerequisite for research progress.

Objective: The aim of the study was to learn about patients' attitudes and expectations regarding secondary use of their clinical data. In a next step, our results can inform the development of an appropriate governance framework for secondary use of clinical data for research purposes.

Methods: A questionnaire was developed to assess the willingness of patients with cancer to provide their clinical data for biomedical research purposes, considering different conditions of data sharing and consent models. The Cancer Registry of the German federal state of Baden-Württemberg recruited a proportionally stratified random sample of patients with cancer and survivors of cancer based on a full census.

Results: In total, 838 participants completed the survey. Approximately all participants (810/838, 96.7%) showed general willingness to make clinical data available for biomedical research purposes; however, they expected certain requirements to be met, such as comparable data protection standards for data use abroad and the possibility to renew consent at regular time intervals. Most participants (620/838, 73.9%) supported data use also by researchers in commercial companies. More than half of the participants (503/838, 60%) were willing to give up control over clinical data in favor of research benefits. Most participants expressed acceptance of the broad consent model (494/838, 58.9%), followed by data use by default (with the option to opt out at any time; 419/838, 50%); specific consent for every study showed the lowest acceptance rate (327/838, 39%). Patients expected physicians to share their data (763/838, 91.1%) and their fellow patients to support secondary use with their clinical data (679/838, 81%).

Conclusions: Although patients' general willingness to make their clinical data available for biomedical research purposes is very high, the willingness of a substantial proportion of patients depends on additional requirements. Taking these perspectives into account is essential for designing trustworthy governance of clinical data reuse and sharing. The willingness to accept the loss of control over clinical data to enhance the benefits of research should be given special consideration.

(J Med Internet Res 2022;24(8):e37665) doi: 10.2196/37665

KEYWORDS

secondary use; consent; data sharing; data access; research benefit and control of data; health data; clinical data; private sector; international data sharing; patient perspective

Introduction

Background

Secondary use of clinical data for biomedical research purposes has great potential for various types of noninterventional, data-driven studies. We define secondary use of clinical data as the collection and reuse of clinical data in data gathering, noninterventional biomedical research, or learning activities; clinical data are collected during and for the purpose of patient care [1]. Research using clinical data has the ethical and efficiency advantages of not requiring additional physical interventions or collection of additional data. Although secondary use aims at improving biomedical knowledge and, in turn, medical care, it does not imply a direct benefit for the patient who has released their data.

The blurring of the boundaries between research and care, as envisaged in concepts of learning health care systems, is currently visible only in few areas [1,2]. The endeavor to merge these different system logics is faced with emerging challenges such as limited utility of specific consent models for research or false expectations regarding their benefits on the part of patients [3]. The goal of this paper was to contribute the patients' perspective to the debate and potential solutions to the current challenges of secondary use of clinical data in the context of learning health care systems.

Previous studies with citizens and patients have already shown that certain aspects seem to be crucial for supporting secondary use, such as who conducts the research (eg, academic or commercial), whether data are transferred to other countries, and what consent model is applied [2-9]. However, owing to varying research designs, for example, by examining different study units, applying different survey instruments, and being conducted in diverse health care systems, these studies, taken together, have heterogeneous results.

Consent is a crucial component of respecting patient autonomy and building trust in health research. However, the specific consent paradigm of clinical trials cannot easily be applied to the secondary use of clinical data because most scientific questions are unknown at the time consent is obtained, that is, when the patient receives care. Newly applied models for secondary use of clinical data, such as broad consent or data use by default (with the option to opt out any time), facilitate research with clinical data, but are criticized from an informational self-determination perspective for offering patients insufficient control over their clinical data. However, previous studies have identified patients' and citizens' openness toward these new models [5,10-12]. Other empirical studies show that, to increase research benefits, participants seem willing to accept the loss of control over their data [13-16]. However, no studies have yet been conducted to assess the acceptance of consent models in light of the trade-off between the control of clinical data and research utility.

Aim

The objective of this study was to assess (1) patients' general willingness and relevant requirements to share pseudonymized clinical data for research purposes, (2) acceptance of different consent models including characteristics of data control and research utility, (3) preferences regarding the setting to provide consent, and (4) general expectations toward data use and other stakeholders.

Methods

Survey Development

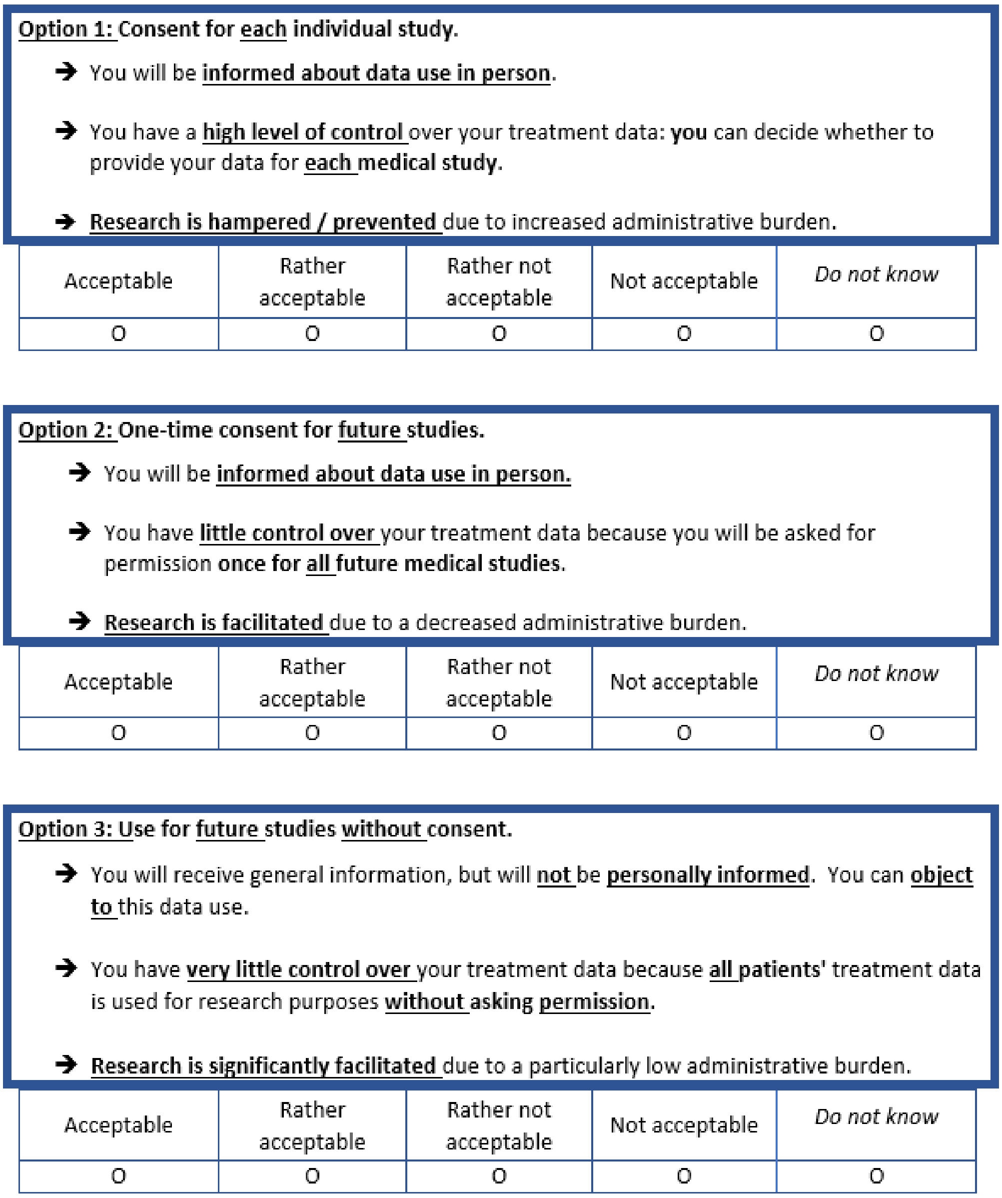
The questionnaire (Multimedia Appendix 1) was based on a review of the relevant scientific literature and a preparatory expert interview study among stakeholder groups engaged or affected by the planned secondary use of clinical data in Germany [17]. In total, 2 representatives for patient interests were included in the expert sample. The questionnaire was developed through several discussion and feedback rounds by the international and interdisciplinary project team, consisting of social scientists; ethicists; legal scholars; and clinicians with expertise in social, ethical, legal, or practical aspects of secondary use of clinical data. To ensure comprehensibility and technical functionality of the questionnaire, cognitive interviews (n=5) with patients with cancer and survivors of cancer who had provided consent were conducted in the pretest phase, resulting in minor adaptions.

To allow participants to develop an informed opinion, the survey included background information about risks and benefits associated with the secondary use of clinical data. The survey consisted of 33 items on the following topics: sociodemographic and disease-related information, expectations and risk perception toward secondary use, willingness to provide clinical data under certain requirements, and acceptability of consent models and procedures. Attitudinal questions were designed as 5-point Likert scale. The survey was approved by the data protection officer of the Heidelberg University Hospital.

Operationalization of Consent Scenarios

In total, 3 vignettes were developed to measure the acceptability of 3 consent scenarios: *specific consent*, *broad consent*, and *data use by default* (with the option to opt out at any time). Acceptance was measured using a 4-point Likert scale (Figure 1).

Figure 1. Display of the 3 consent scenarios in the questionnaire (English translation).



Option 1: Consent for each individual study.

- ➔ You will be **informed about data use in person**.
- ➔ You have a **high level of control** over your treatment data: **you** can decide whether to provide your data for **each medical study**.
- ➔ **Research is hampered / prevented** due to increased administrative burden.

Acceptable	Rather acceptable	Rather not acceptable	Not acceptable	*Do not know*
○	○	○	○	○

Option 2: One-time consent for future studies.

- ➔ You will be **informed about data use in person.**
- ➔ You have **little control over** your treatment data because you will be asked for permission **once for all future medical studies**.
- ➔ **Research is facilitated** due to a decreased administrative burden.

Acceptable	Rather acceptable	Rather not acceptable	Not acceptable	*Do not know*
○	○	○	○	○

Option 3: Use for future studies without consent.

- ➔ You will receive general information, but will **not** be **personally informed**. You can **object to** this data use.
- ➔ You have **very little control over** your treatment data because **all patients'** treatment data is used for research purposes **without asking permission**.
- ➔ **Research is significantly facilitated** due to a particularly low administrative burden.

Acceptable	Rather acceptable	Rather not acceptable	Not acceptable	*Do not know*
○	○	○	○	○

Previous studies have reported that participants made a trade-off between research utility and data control [13]. Hence, 3 consent scenarios were designed with information about research utility and control over data. In the process of operationalization, we further reduced the complexity of the theoretical concept to ensure good comprehensibility of the survey material:

1. By *specific consent*, we understand that consent is provided for each individual study (option 1 in Figure 1), as currently performed in clinical trials. Consistent with our preliminary studies [1,17], we inform about high degree of control over the secondary use of clinical data and low research benefit owing to the administrative burden on researchers.
2. In the case of the *broad consent* scenario, 1-time consent is provided for future medical studies with clinical data; moderate control and research utility are presumed (option 2 in Figure 1).
   This vignette refers to the implementation of a broad consent process for the German Medical Informatics Initiative; the development of a unified template for broad consent was accompanied by the German Working Group of Research Ethics Committees [18]. In practice, this model involves safeguards such as the review of each individual research project by a research ethics committee and data access committees, organizational measures to protect patient data, and comprehensive information for patients [18]. To ensure comprehensibility, the details of these safeguards are not provided to the participants of this study.
3. *Data use by default* is use of data for secondary research by default (comparable with Denmark or Estonia) without individual informed consent process, but with the possibility to opt out at any time. This scenario is associated with low degree of data control for patients and facilitation of research as no individual consent needs to be obtained (option 3 in Figure 1).
   Regarding law, the European Union (EU) data protection regulation provides some scope for this scenario of data use based on a legal basis other than informed consent if

> the potential research benefit clearly outweighs the right to informational self-determination (Art. 9, Paragraph 2, lit. j [19]). Compensating efforts such as ambitious security and privacy measures and extensive general public education about data use and data governance are likely to be ethically and legally necessary. To ensure comprehensibility, the details of these safeguards are not provided to the participants of this study.

Sampling and Recruitment

The Cancer Registry of the German federal state of Baden-Württemberg sent postal invitations to a random sample of patients with cancer and survivors of cancer, proportionally stratified by age and gender, requesting study participation (n=4219). The sample frame consisted of all registered patients in Baden-Württemberg, Germany, with a diagnosed tumor disease who were aged ≥18 years. Participants had the option of either completing an anonymous and self-administered web-based survey (the hyperlink was provided in the cover letter) or returning an envelope by mail, consenting that their address may be forwarded to the research group to subsequently receive a paper-and-pencil questionnaire. Survey instruments were adapted to the requirements of a mixed-mode survey [20].

Individuals who completed the survey were not compensated.

Data collection occurred from May 2021 to July 2021.

Analysis

Descriptive statistics were used to express the categorical variables as counts and percentages. Differences in proportions were assessed for statistical significance ($P<.05$) using chi-square tests. The 2-tailed Pearson correlation coefficients were computed. All analyses were performed using SPSS (version 28; IBM Corp).

Ethics Approval

The study obtained ethics approval from the University of Heidelberg's research ethics committee (reference number S-361/2018). Informed consent was obtained from the individuals who participated in the study pretest measurement and the written survey.

Results

Demographics of Participants

Of the 4155 patients with cancer approached by the Cancer Registry Baden-Württemberg, 838 (20.17%) participants completed the survey. Approximately half of the participants who answered the respective question were women (389/820, 47.4%; Table 1). Of 832 participants, 390 (46.9%) participants were aged between 60 and 74 years, and of 826 participants, 541 (65.5%) participants were retired. In total, 29.8% (247/830) of the participants had a university degree. The most common types of cancer were breast cancer (204/826, 24.7%), prostate cancer (187/826, 22.6%), and gastrointestinal cancer (79/826, 9.6%). The distribution of age, gender, and cancer entity mirrored that of the general distribution of patients with cancer in the Cancer Registry Baden-Württemberg, with minor deviation.

Table 1. Demographics of participants.

Characteristics	Values, n (%)
Gender (n=820)	
Women	389 (47.4)
Men	431 (52.6)
Age groups (years; n=832)	
18-59	186 (22.4)
60-74	390 (46.9)
≥75	256 (30.8)
Highest educational degree (n=830)	
Elementary school diploma	84 (10.1)
Secondary school diploma	398 (47.9)
Qualification for university entrance	97 (11.7)
University degree	247 (29.8)
No school diploma	4 (0.5)
Employment status (n=826)	
Employed or self-employed	219 (26.5)
Not employed owing to health reasons	45 (5.4)
Retired	541 (65.5)
Not employed owing to other reasons	21 (2.5)
Type of cancer (n=826)	
Breast	204 (24.7)
Prostate	187 (22.6)
Gastrointestinal	79 (9.6)
Skin cancer	63 (7.6)
Non-Hodgkin lymphoma	39 (4.7)
Lung	31 (3.8)
Leukemia	22 (2.7)
Kidney	22 (2.7)
Head and neck	22 (2.7)
Uterine or endometrial	21 (2.5)
Urinary bladder	18 (2.2)
Stomach	16 (1.9)
Pancreas	9 (1.1)
Other	93 (11.3)

General Willingness to Provide Clinical Data for Biomedical Research Purposes and Requirements for Data Provision

Most participants indicated that they are generally willing to make their clinical data available either without restrictions (527/838, 62.9%) or under certain conditions (283/838, 33.8%). Only 0.7% (6/838) of the participants generally refused to provide clinical data.

Then, the participants who indicated general willingness were asked about certain requirements under which they would provide their clinical data. When asked about the general requirements they deemed relevant, most participants stated the highest possible data security standards (482/838, 57.5%), followed by use of their data for as many research projects as possible (254/838, 30.3%), and being informed about the most important research results (208/838, 24.8%; Multimedia Appendix 2).

Most participants (591/832, 70.5%) stated that they would support research with their data in countries with high level of data protection comparable with German standards; 17.9% (149/832) of the participants stated that they would restrict data use to domestic research projects; and 8.8% (73/832) of the participants agreed to support international projects, independent of the level of data protection (Multimedia Appendix 3).

When asked how long their initial consent should be valid, 38.5% (320/832) of the participants set no time limit and approximately half of the participants demanded to renew consent either after 3 years (181/832, 21.8%), 10 years (227/832, 27.3%), or 30 years (10/832, 1.2%), respectively. In total, 10.2% (85/832) of the participants favored renewal of consent each time their data are used for specific research projects (Multimedia Appendix 4).

A large proportion of participants (532/832, 63.4%) said that they would grant access to researchers, independent of their affiliation; however, 22.7% (189/832) of them did not want to share their data with researchers at for-profit companies that conduct medical research (Multimedia Appendix 5). Only a small proportion opposed the secondary use of their clinical data by their physicians (56/832, 6.7%) or researchers at universities and university hospitals (48/832, 5.8%).

Acceptance of Consent Models

The questionnaire provided information about 3 consent models that correspond to specific consent, broad consent, and data use by default (with the option to opt out at any time), including the trade-offs of each model between control over clinical data and the facilitation of medical research (Table 2). For each consent model, the participants rated the level of acceptance on a 4-point Likert scale. Each of the 3 consent models showed a medium degree of acceptance with significant mean differences. Of the 838 participants, 491 (58.6%) accepted the broad consent model, 421 (50.2%) accepted data use by default (with the option to opt out at any time), and 323 (38.5%) accepted the specific consent model. Of the 323 participants accepting the specific consent model, 102 (31.6%) did not accept any other model (102/838, 12.2% of the total sample). Sociodemographic characteristics were not significant, except for older participants being more likely to accept data use by default (Pearson coefficient, 2-tailed: r=0.138; P<.001).

Table 2. Acceptance rates of 3 consent models: broad consent, data use by default, and specific consent (N=838)[a].

Model	Description	Accepted, n (%)	Not accepted, n (%)	Do not know or not answered, n (%)
Broad consent	One-time consent for future studies, informed in person, low level of control, and research is facilitated	491 (58.6)	230 (27.4)	117 (13.9)
Data use by default	Use for future studies without consent process, not personally informed, very low level of control, and research is significantly facilitated	421 (50.2)	347 (41.4)	70 (8.4)
Specific consent	Consent for each study, informed in person, high level of control, and research is hampered	323 (38.5)	372 (44.4)	143 (17.1)

[a] Acceptance was measured using a 4-point scale; results were collapsed into 2 groups (*not acceptable*: not acceptable and rather not acceptable; *acceptable*: acceptable and rather acceptable).

Preferences Regarding the Setting for Providing Consent

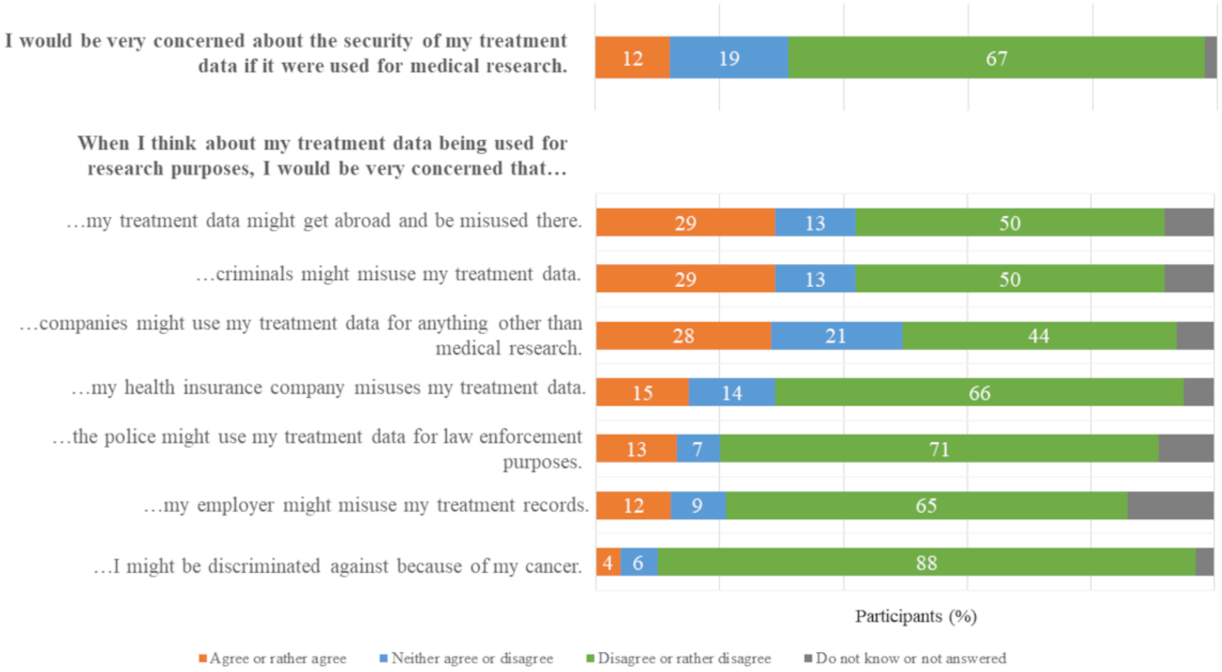
Participants were asked about the most appropriate setting for providing consent for the secondary use of their clinical data for research purposes. Most of them preferred to decide at their general practitioner's practice (528/838, 63%), and a small proportion of participants preferred to decide during the admission to a hospital (174/838, 20.8%; Multimedia Appendix 6).

When asked about preferred information formats, most participants selected a brief written summary of key points in easy-to-understand language to learn more about secondary use (616/838, 73.5%), followed by face-to-face consultation with physicians (347/838, 41.4%; Multimedia Appendix 7). Participants were asked about who should decide about data access and use by individual research projects: most participants (393/838, 46.9%) favored committees with experts in which the opinion of patients is represented, for example, by patient representatives, whereas a small proportion of participants preferred to leave the decision to an expert committee (without patient representation; 185/838, 22.1%) or to decide for themselves (200/838, 23.9%; Multimedia Appendix 8).

Concerns in the Event of Data Use

A small proportion of the participants (99/838, 11.8%) showed major general concerns regarding their clinical data being used for research purposes (Figure 2). Then, all participants were asked about more specific concerns: the largest proportion of participants were worried about the data being misused in countries other than Germany (246/838, 29.4%), data being misused by criminals (244/838, 29.1%), and data being used by companies for something other than medical research (235/838, 28%). Concerns about participants being discriminated against because of cancer were very low (32/838, 3.8%).

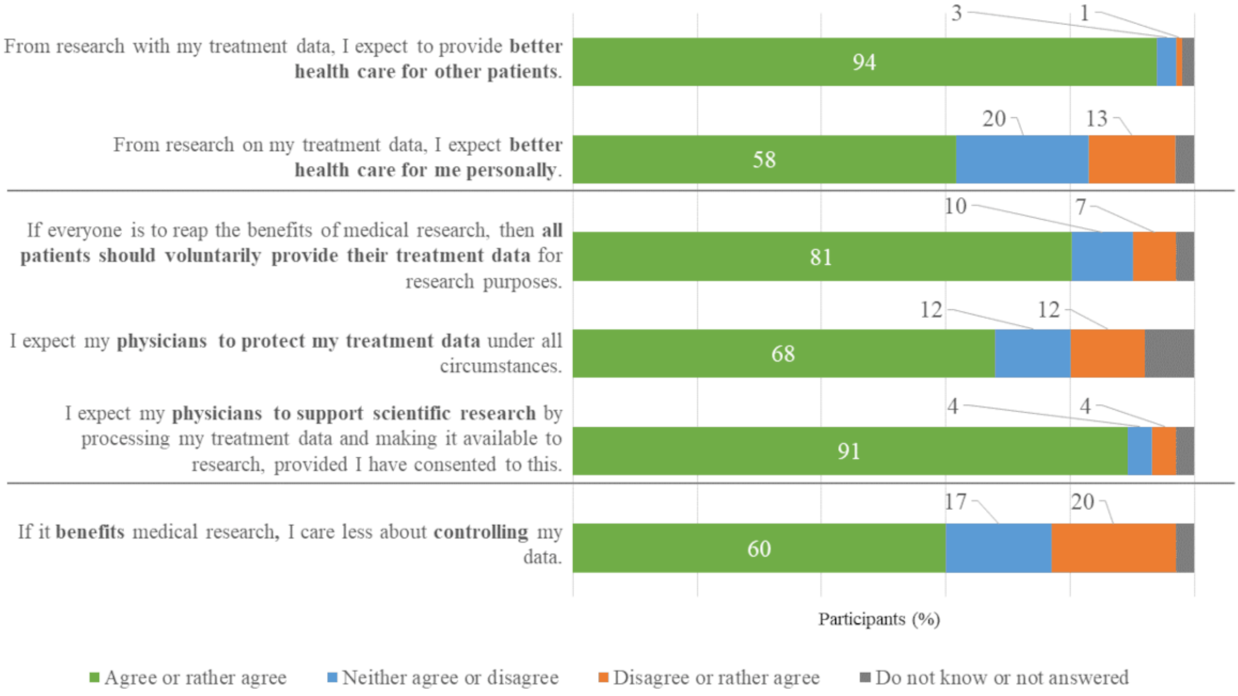
Figure 2. Concerns in the event of data use (N=838).



Expectations Toward Benefits, Other Patients, and Physicians

Approximately all participants (788/838, 94%) expected a benefit for other patients from making their clinical data available for research purposes (Figure 3). More than half of the participants (482/838, 57.5%) mistakenly expected a personal benefit, even though the explanatory text explicitly stated the opposite. Of the 838 participants, 676 (80.7%) participants supported the claim that all patients should voluntarily make their clinical data available for research purposes. In total, 68.3% (572/838) of the participants expected their physicians to protect the participants’ clinical data in all circumstances, and approximately all participants (758/838, 90.5%) expected their physicians to support research, if consent was provided, by making their patients’ clinical data available for research.

Figure 3. Expectations toward physicians and other patients (N=838).



Discussion

Principal Findings

Information about the requirements under which patients would make their treatment data available for research is important for any form of policy that regulates the secondary use of such data. This paper provides the results of a representative sample of German patients with cancer on general willingness and decisive requirements for sharing their data for research purposes and on the acceptance of consent models and expectations toward relevant stakeholders. The following are the main findings. First, we found an unprecedentedly high general willingness (810/838, 96.7%) to make clinical data available even after being informed about the potential risks of secondary use; however, relevant requirements included the following: ensuring a high level of data security, comparable data protection standards for data use abroad, and renewed consent at regular time intervals. Second, in contrast to previous studies, three-fourths of respondents (620/838, 73.9%) supported data use also by researchers in commercial companies. Third, the highest acceptance rate was found for a broad consent model (494/838, 58.9%), followed by data use by default (419/838, 50%); and specific consent for every study (327/838, 39%). Fourth, high expectations for physicians and fellow patients to support data sharing for research purposes were found.

To the best of our knowledge, this is the first representative study on attitudes toward the secondary use of clinical data and acceptance of consent models in combination with characteristics of data control and research utility.

High General Willingness to Provide Clinical Data

Overview

An important finding of our study was the high willingness of patients with cancer to make their clinical data available for research purposes (810/838, 96.7%), either without any restrictions (527/838, 62.9%) or under certain conditions (283/838, 33.8%). Only 0.7% (6/838) of the participants generally refused to provide clinical data. A population-representative study in the United States found low proportions of general willingness (76%) [21], similar to representative studies in Germany in the contexts of the COVID-19 pandemic (65%) [22] and medical data including genetic data (56%) [13]. These different results suggest that patients with cancer are more willing to provide clinical data for medical research because they may either have benefited or hope to benefit from research. As potential beneficiaries of past studies, they may also feel greater responsibility than citizens and other patient groups to support research to help future generations of patients [11,23-25]. Although patients with cancer are not representative of all patients, we assume that they can hint well at the attitude of other patient groups with severe or rare diseases, such as leukodystrophies [26]. A study conducted in the United States shows slightly lower willingness among patients with cancer and survivors of cancer (71%) [27] than among the general population (76%) [21], which may point toward country-specific factors in the context of health systems and trust in institutions.

General Requirements: Data Security, Maximizing Data Use, and Transparency

The most relevant general requirements for supporting the secondary use of clinical data for research were high data security (486/838, 57.9%), maximizing data use (251/838, 29.9%), and information about research results that made use of patients' clinical data (210/838, 25.1%). These findings indicate the relevance of the ability of data governance to protect clinical data, maximize accessibility (and usability) of data for research, and report transparently on the results of data use. These findings are largely consistent with previous literature that describes secure data use, public benefits through effective use by researchers, and transparency as important requirements for data sharing [3,8,14]. It may well be that participants value the reporting of results as an act of recognition and reciprocity. Suggestions for future set up of governance for secondary use of data to respond to the abovementioned requirements include appropriate safeguards to protect patient data; high degree of transparency regarding data use and benefits to society; and technical, organizational, and legal data infrastructure that enables researchers to maximize research benefits. Involving patients to better understand their concrete needs in designing these requirements for secondary use seems advisable [28].

Data Transfer Only to Countries With Comparable Data Protection Standard

Most participants stated that they would restrict their data to research in countries with data protection standards comparable with those in Germany (737/838, 87.9%), and a small minority of the participants was willing to provide data to other countries (75/838, 8.9%). This resonates with another German study with outpatients who generally support data donation in favor of public research institutions in EU countries with similar data protection standards (92%); only a minority of the participants approved data access to countries outside the EU (24%), which is a large share compared with our findings [6]. The high relevance of this aspect is consistent with studies of Canadian citizens [8,9]. However, further studies are needed to explore the exact kinds of misuse that make people fearful about international data transfers. Our study suggests that comparable data protection standards are a decisive requirement for patients. A suggestion to address this need is that policy makers and data initiatives explain well to patients what the additional benefit of multinational research is, what the specific risks are (eg, foreign government access and less ability to enforce rights), and how risks to data protection in these countries are mitigated. They are well advised to give patients the choice of whether to consent to data transfer to countries with low data protection standards.

Most Participants Support Data Use by Corporate Researchers

Low willingness of citizens to share data with the private industry has been reported in several studies [3,4]. This finding poses challenges to the biomedical research landscape, as many studies are conducted by companies or in cooperation with companies. In contrast, our results show that approximately three-fourths of the participants (620/838, 73.9%) were willing to make their clinical data available to company researchers.

This is a much higher acceptance than in studies with German citizens [5] and outpatients [6], which reported that only a minority of those participants who agreed to data donation were willing to provide data to the industry (17% and 29%, respectively). A cross-country study found particularly low support for medical and genetic data sharing with for-profit researchers among German participants (22% compared with 32% on average across all countries) [29]. We hypothesize that willingness to share data with company researchers may change owing to experiences with a severe illness: patients with cancer may develop strong awareness of contributions by corporate researchers, possibly based on their experiences during their therapy. In addition, our questionnaire item included a brief explanation of the relevant contribution of industry to medical research and of industry as an important collaborator with public research institutions. We suppose the explanation increased the participants' understanding and willingness to provide clinical data to the industry, which is consistent with a study examining public attitudes toward commercial data access, in which provision of information and deliberative methods increased willingness to share data [15]. In addition, our findings indicate that low willingness to share data with corporate researchers can be addressed through collaboration with public research institutions in public-private partnerships.

Renewed Consent Within Certain Time Intervals

The participants' stance was divided on the duration of data use after initial consent is provided. Most participants (408/838, 48.7%) preferred to renew consent for broad research use after a period of 3 or 10 years. Only approximately one-third of the participants (243/838, 28.9%) preferred 1-time consent with unlimited duration of consent validity. In contrast, in a representative study of German citizens, more than half of the participants favored unlimited validity of consent (56%), and a minority favored consent validity of 5 years (17%) [5]. Our reported relatively high proportion of participants preferring renewed consent may have resulted owing to the following reasons. First, patients with cancer experience changing health conditions, leading to a subjective sensitivity to release clinical data without time limit. Second, our questionnaire explicitly mentioned risks of data release, possibly reducing the approval of unlimited data use. Third, the abovementioned study among German citizens asked for unlimited use for "data donation," which can be understood as irrevocable by definition. To address this potential need for patients to renew consent, further studies should investigate the preferences using neutral wording.

Broad Consent and Data Use by Default Was More Accepted Than Specific Consent—Research Benefits Partially Outweigh Loss of Control

Overview

Participants were presented with general information about 3 consent models (specific consent for every study; broad consent; and data use by default, with the option to opt out at any time). Specific consent is related to maximum informational control for patients, but less utility for research projects, whereas data use by default is associated with less informational control, but maximum utility for research projects. The broad consent model features moderate control and research utility (Table 2). The opportunity of being personally given information by health personnel is not available in the case of data use by default. Participants rated the level of acceptance for each consent model. The broad consent model received the highest acceptance rate (491/838, 58.6%), followed by data use by default (421/838, 50.2%) and the specific consent model with only a moderate acceptance rate (323/838, 38.5%). The relatively high acceptance rate for the broad consent model is consistent with the results of previous studies. Different study designs and minor deviations regarding the definition of consent procedures apply; therefore, comparisons should be considered cautiously. In total, 2 studies with a German patient sample and a large sample of Dutch patients found even higher acceptance rates in the context of health care–embedded biobanking and data donation (92%-93%) [5,10]. An earlier study of German patients (87%) [11] and a study of a smaller sample of US citizens (96%) [12] showed similar results. Our acceptance rates for each of the presented consent models were lower than those in other studies. This may be a consequence of the choice among 3 different models, rather than only 1, as presented in other studies. The low acceptance rates may also result from a trade-off decision between support for research and control over one's clinical data. Previous studies have described this trade-off between control and research benefits as a relevant influencing factor in decision-making [12-16]. Accordingly, in our study, most participants (520/838, 62.1%) agreed to give up control if it increased the benefits of research. This finding is significant because most participants (804/838, 95.9%) believe in the benefits of secondary use for other patients. Evidence from other studies [11,23-25] and our findings not only suggest that research benefits partially outweigh the loss of control but also that they are a critical motivational aspect of making data available for research.

As none of the models achieved wide-ranging acceptance in our study, it is worth discussing whether a meta-consent model that allows participants to choose their preferred consent variants [30] accounts best for individual ways of balancing control and research benefits regarding consent models.

Preferred Framework Conditions for Providing Consent and Data Release

When asked for consent, participants expected brief and understandable written information (616/838, 73.5%) about data use and preferred their primary care physician as a venue for informed consent (528/838, 63%) over providing consent upon hospital admission (174/838, 20.8%). This finding is underpinned by a qualitative study in which support by health care professionals was seen as an important facilitator [8].

Our findings indicate that, regarding place and time (ie, where and when patients are informed and asked for consent), consent in the clinical context is preferred over consent before becoming a patient. This is consistent with the finding that patients prefer providing consent at hospitals (64%-76%) over providing consent outside the clinic [6]. However, another study concluded that the decision about making data available for research should be separated from the clinical context and anchored in everyday life [31]. Owing to possible age and disease effects, further studies should investigate the differences between the general population's and patients' acceptance.

When asked who should decide on data release when individual research projects apply for using participants' clinical data after having personally released their treatment data for research purposes, approximately half of the participants (394/838, 47%) preferred a committee with experts and patient representatives over a committee with experts only or deciding for themselves. A suggestion to address this need is to involve patients in data access committees.

Low Concerns and High Expectations

Low Level of Concern in General and About Discrimination

In our study, the proportion of participants who were concerned about the use of clinical data (101/838, 12.1%) was considerably lower than the findings of 2 surveys conducted in Australia among citizens (24%-25%) [32] and patients (24%) [33]. A study conducted in the United States found that privacy concerns had the strongest influence on individuals' intentions to provide clinical data [34]. This discrepancy may be attributable to country-specific differences regarding trust in health care and government institutions [3] and the lack of experience with extensive data leaks or the misuse of clinical data in Germany. Participants' concerns about discrimination owing to their cancer were very low (34/838, 4.1%).

False Expectations of Personal Benefit

Most participants (486/838, 57.9%) incorrectly expected personal benefits from making their clinical data available for research purposes—even though the wording of the questionnaire had been adjusted during the pretest phase. Another study found that more than one-fourth of German patients hoped for personal benefit (28%) after being asked for consent for secondary use of clinical data and biomaterial collected during routine care [11]. Owing to the severity of the disease, patients with cancer may be particularly prone to this false expectation of personal benefit from research with their health data, which is comparable with therapeutic misconception [35] in clinical trials. The study showed that the proportion of those holding false expectation decreased considerably after the modification of consent information material (12%). To reduce the risk of false expectations, particularly in vulnerable groups such as patients with severe illnesses, careful education about the unlikelihood of direct benefits from making their clinical data available for research purposes is needed.

High Expectations of Other Patients and Physicians

Our results indicate a clear expectation toward fellow patients (696/838, 83.1%) to support medical research with clinical data, which is consistent with a study conducted in Germany among outpatients (80%-90%) [6]. Interestingly, more participants expected their physicians to share clinical data for research (754/838, 89.9%) than to protect their clinical data under all circumstances (570/838, 68%). This is the first study to investigate the expectations toward physicians.

Limitations

The recruited sample is largely representative of the population of patients with cancer in the federal state of Baden-Württemberg in age, gender, and cancer entity. However, we found that the educational level in our sample was higher than that of the corresponding age cohorts of the German population [36]. The educational level of the German population presumably applies to the group of patients with cancer and survivors of cancer. Owing to the topic of the survey, we suspect a self-selection bias correlating with high educational level. According to a study in the context of genetic research and biobanking, high educational level positively correlates with willingness to provide data [24]; consequently, our results may overestimate willingness to provide clinical data. In addition, false expectations of personal benefits from providing data for secondary use may have increased the participants' willingness to share clinical data.

A considerable proportion of participants who had previously agreed to hypothetically make their clinical data available *without restrictions*, favored restricted use of their clinical data when asked about specific requirements such as data user, duration, and data use in other countries (Multimedia Appendix 9). We assume that the participants have not yet formed a strong opinion about sharing their clinical data. Hence, the general willingness to provide clinical data seems to measure an overall attitude toward secondary use, rather than the actual willingness to provide clinical data without restrictions for research purposes.

Conclusions

Our study shows very high general willingness of patients with cancer to make their clinical data available for biomedical research purposes. However, the willingness to provide clinical data may be overstated owing to the above-average educational level of the respondents. For a considerable proportion of patients with cancer, willingness depends on certain requirements. In addition to the basic prerequisite of high level of data security and transparency in the use of the data, most patients shared the view that the data must not be used in countries with low data protection standards and that they should have the possibility to renew consent. In contrast to previous studies, the exclusion of use of data for private sector studies is not a requirement for most participants.

High willingness on the part of patients to accept loss of control over clinical data in favor of research benefits and request to maximize accessibility (and usability) of data for research were found. This is consistent with the acceptance of more research-friendly and low-control models, namely the broad consent model, followed by data use by default (with the option to opt out at any time). The striving for maximizing data use is also reflected by patients' expectations toward physicians and other patients to support secondary use.

Policy makers are well advised to account for patients' views when designing and implementing secondary use, with the aim to contribute to a socially legitimized culture of data sharing.

Acknowledgments

This paper was developed within the framework of the project, *Learning from Clinical Data*, funded by the Deutsche Forschungsgemeinschaft (German Research Foundation)—406103282. The project funding was awarded to ECW. The funders did not engage in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the paper for publication. No additional external funding was received for this study. The authors would like to thank the Cancer Registry of Baden-Württemberg for supporting data collection. The authors also thank their project partners, Kai Cornelius, Markus Spitz (University of Heidelberg, Germany), and Adrian Thorogood (University of Luxemburg), for conceptual counseling. The authors acknowledge Hellen Kachler and Niamh Sulzbach (both from the National Center for Tumor Diseases, Heidelberg University Hospital, Germany) for their assistance in researching the literature, data entry, and linguistic correction of the manuscript. The authors would also like to express their gratitude to Anne Müller, who assisted them in the recruitment of patients for the pretest phase, and to all the patients with cancer and survivors of cancer who participated in the study.

Data Availability

The data sets generated and analyzed in this study are available on the web [37].

Authors' Contributions

All the authors contributed to the conception and design of the study. AK and KM were involved in material preparation and data collection and analysis. ECW and KM supervised the study and supported data interpretation. AK wrote the first draft of the manuscript, and all the authors commented on the previous versions of the manuscript. All the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire.
[PDF File (Adobe PDF File), 167 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Participants' requirements that are to be met to provide clinical data.
[DOCX File , 14 KB-Multimedia Appendix 2]

Multimedia Appendix 3

The acceptance of use of clinical data for biomedical research purposes in other countries.
[DOCX File , 14 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Preferred duration of use of clinical data for biomedical research purposes.
[DOCX File , 13 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Participants' restrictions for the provision of their clinical data regarding different researcher groups.
[DOCX File , 13 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Preferred context of providing informed consent.
[DOCX File , 13 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Ways of obtaining information.
[DOCX File , 13 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Decision on data release for individual research projects.
[DOCX File , 13 KB-Multimedia Appendix 8]

Multimedia Appendix 9

Conditional data release of participants initially stating to provide data without restriction.
[DOCX File , 13 KB-Multimedia Appendix 9]

References

1. Jungkunz M, Köngeter A, Mehlis K, Winkler EC, Schickhardt C. Secondary use of clinical data in data-gathering, non-interventional research or learning activities: definition, types, and a framework for risk assessment. J Med Internet Res 2021 Jun 08;23(6):e26631 [FREE Full text] [doi: 10.2196/26631] [Medline: 34100760]
2. Braunack-Mayer A, Fabrianesi B, Street J, O'Shaughnessy P, Carter SM, Engelen L, et al. Sharing government health data with the private sector: community attitudes survey. J Med Internet Res 2021 Oct 01;23(10):e24200 [FREE Full text] [doi: 10.2196/24200] [Medline: 34596573]
3. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan;48(1):3-13 [FREE Full text] [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
4. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
5. Richter G, Borzikowsky C, Lesch W, Semler SC, Bunnik EM, Buyx A, et al. Secondary research use of personal medical data: attitudes from patient and population surveys in The Netherlands and Germany. Eur J Hum Genet 2021 Mar;29(3):495-502 [FREE Full text] [doi: 10.1038/s41431-020-00735-3] [Medline: 33005018]
6. Richter G, Borzikowsky C, Hoyer BF, Laudes M, Krawczak M. Secondary research use of personal medical data: patient attitudes towards data donation. BMC Med Ethics 2021 Dec 15;22(1):164 [FREE Full text] [doi: 10.1186/s12910-021-00728-x] [Medline: 34911502]
7. Shah N, Coathup V, Teare H, Forgie I, Giordano GN, Hansen TH, et al. Motivations for data sharing-views of research participants from four European countries: a DIRECT study. Eur J Hum Genet 2019 May;27(5):721-729 [FREE Full text] [doi: 10.1038/s41431-019-0344-2] [Medline: 30700834]
8. Cumyn A, Barton A, Dault R, Safa N, Cloutier A, Ethier J. Meta-consent for the secondary use of health data within a learning health system: a qualitative study of the public's perspective. BMC Med Ethics 2021 Jun 29;22(1):81 [FREE Full text] [doi: 10.1186/s12910-021-00647-x] [Medline: 34187453]
9. McCormick N, Hamilton CB, Koehn CL, English K, Stordy A, Li LC. Canadians' views on the use of routinely collected data in health research: a patient-oriented cross-sectional survey. CMAJ Open 2019;7(2):E203-E209 [FREE Full text] [doi: 10.9778/cmajo.20180105] [Medline: 30948649]
10. Richter G, Borzikowsky C, Lieb W, Schreiber S, Krawczak M, Buyx A. Patient views on research use of clinical data without consent: legal, but also acceptable? Eur J Hum Genet 2019 Jun;27(6):841-847 [FREE Full text] [doi: 10.1038/s41431-019-0340-6] [Medline: 30683927]
11. Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A. Broad consent for health care-embedded biobanking: understanding and reasons to donate in a large patient sample. Genet Med 2018 Jan;20(1):76-82 [FREE Full text] [doi: 10.1038/gim.2017.82] [Medline: 28640237]
12. Hammack-Aviran CM, Brelsford KM, McKenna KC, Graham RD, Lampron ZM, Beskow LM. Research use of electronic health records: patients' views on alternative approaches to permission. AJOB Empir Bioeth 2020;11(3):172-186 [FREE Full text] [doi: 10.1080/23294515.2020.1755383] [Medline: 32338567]
13. Voigt TH, Holtz V, Niemiec E, Howard HC, Middleton A, Prainsack B. Willingness to donate genomic and other medical data: results from Germany. Eur J Hum Genet 2020 Aug;28(8):1000-1009 [FREE Full text] [doi: 10.1038/s41431-020-0611-2] [Medline: 32238912]
14. Tully MP, Bozentko K, Clement S, Hunn A, Hassan L, Norris R, et al. Investigating the extent to which patients should control access to patient records for research: a deliberative process using citizens' juries. J Med Internet Res 2018 Mar 28;20(3):e112 [FREE Full text] [doi: 10.2196/jmir.7763] [Medline: 29592847]
15. Ipsos M. The One-Way Mirror: public attitudes to commercial access to health data. Wellcome Trust. 2017. URL: https://wellcome.figshare.com/articles/journal_contribution/The_One-Way_Mirror_Public_attitudes_to_commercial_access_to_health_data/5616448/1 [accessed 2022-07-21]
16. Oliver JM, Slashinski MJ, Wang T, Kelly PA, Hilsenbeck SG, McGuire AL. Balancing the risks and benefits of genomic data sharing: genome research participants' perspectives. Public Health Genomics 2012;15(2):106-114 [FREE Full text] [doi: 10.1159/000334718] [Medline: 22213783]

17. Köngeter A, Jungkunz M, Winkler E, Schickhardt C, Mehlis K. Sekundärnutzung klinischer Daten aus der Patientenversorgung für Forschungszwecke – Eine qualitative Interviewstudie zu Nutzen- und Risikopotenzialen aus Sicht von Expertinnen und Experten für den deutschen Forschungskontext. In: Datenreiche Medizin und das Problem der Einwilligung. Berlin, Heidelberg: Springer; 2021.
18. Zenker S, Strech D, Ihrig K, Jahns R, Müller G, Schickhardt C, et al. Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: towards a new German national standard. J Biomed Inform 2022 Jul;131:104096. [doi: 10.1016/j.jbi.2022.104096] [Medline: 35643273]
19. Regulation (EU) 2016/679 of the European parliament and of the council. Official Journal of the European Union. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN [accessed 2022-08-03]
20. Stadtmüller S, Beuthner C, Silber H. Mixed-Mode Surveys (Version 1.0). GESIS - Leibniz-Institut für Sozialwissenschaften. 2020. URL: https://www.ssoar.info/ssoar/bitstream/handle/document/71311/ssoar-2020-stadtmuller_et_al-Mixed-Mode_Surveys_Version_10.pdf?sequence=1&isAllowed=y&lnkname=ssoar-2020-stadtmuller_et_al-Mixed-Mode_Surveys_Version_10.pdf [accessed 2022-07-21]
21. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. J Am Med Inform Assoc 2015 Jul;22(4):821-830 [FREE Full text] [doi: 10.1093/jamia/ocv014] [Medline: 25829461]
22. Pulscheck Deutschland zu COVID-19 und Datenspende. Data4Life. URL: https://www.data4life.care/de/corona/umfrage/ [accessed 2022-03-03]
23. Nobile H, Bergmann MM, Moldenhauer J, Borry P. Participants' accounts on their decision to join a cohort study with an attached biobank: a qualitative content analysis study within two German studies. J Empir Res Hum Res Ethics 2016 Jul;11(3):237-249. [doi: 10.1177/1556264616657463] [Medline: 27381010]
24. Kettis-Lindblad A, Ring L, Viberth E, Hansson MG. Genetic research and donation of tissue samples to biobanks. What do potential sample donors in the Swedish general public think? Eur J Public Health 2006 Aug;16(4):433-440. [doi: 10.1093/eurpub/cki198] [Medline: 16207726]
25. Aitken M, Porteous C, Creamer E, Cunningham-Burley S. Who benefits and how? Public expectations of public benefits from data-intensive health research. Big Data Soc 2018 Dec 06;5(2):205395171881672. [doi: 10.1177/2053951718816724]
26. Darquy S, Moutel G, Lapointe A, D'Audiffret D, Champagnat J, Guerroui S, et al. Patient/family views on data sharing in rare diseases: study in the European LeukoTreat project. Eur J Hum Genet 2016 Mar 17;24(3):338-343 [FREE Full text] [doi: 10.1038/ejhg.2015.115] [Medline: 26081642]
27. Franklin EF, Nichols HM, House L, Buzaglo J, Thiboldeaux K. Cancer patient perspectives on sharing of medical records and mobile device data for research purposes. J Patient Exp 2020 Dec;7(6):1115-1121 [FREE Full text] [doi: 10.1177/2374373520923837] [Medline: 33457553]
28. Muller SH, Kalkman S, van Thiel GJ, Mostert M, van Delden JJ. The social licence for data-intensive health research: towards co-creation, public value and trust. BMC Med Ethics 2021 Aug 10;22(1):110 [FREE Full text] [doi: 10.1186/s12910-021-00677-5] [Medline: 34376204]
29. Middleton A, Milne R, Almarri MA, Anwer S, Atutornu J, Baranova EE, et al. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data? Am J Hum Genet 2020 Oct 01;107(4):743-752 [FREE Full text] [doi: 10.1016/j.ajhg.2020.08.023] [Medline: 32946764]
30. Ploug T, Holm S. Eliciting meta consent for future secondary research use of health data using a smartphone application - a proof of concept study in the Danish population. BMC Med Ethics 2017 Aug 15;18(1):51 [FREE Full text] [doi: 10.1186/s12910-017-0209-6] [Medline: 28810914]
31. Lesch W, Richter G, Semler S. Daten teilen für die Forschung: Einstellungen und Perspektiven zur Datenspende in Deutschland. In: Datenreiche Medizin und das Problem der Einwilligung. Berlin, Heidelberg: Springer; 2022.
32. King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. Int J Med Inform 2012 Apr;81(4):279-289. [doi: 10.1016/j.ijmedinf.2012.01.005] [Medline: 22306206]
33. Krahe M, Milligan E, Reilly S. Personal health information in research: perceived risk, trustworthiness and opinions from patients attending a tertiary healthcare facility. J Biomed Inform 2019 Jul;95:103222 [FREE Full text] [doi: 10.1016/j.jbi.2019.103222] [Medline: 31176040]
34. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
35. Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: consent to research and the therapeutic misconception. Hastings Cent Rep 1987 Apr;17(2):20-24. [Medline: 3294743]
36. Bildungsstand: Verteilung der Bevölkerung in Deutschland nach höchstem Schulabschluss. Statista. URL: https://de.statista.com/statistik/daten/studie/197269/umfrage/allgemeiner-bildungsstand-der-bevoelkerung-in-deutschland-nach-dem-alter/#professional [accessed 2022-03-03]

37. Köngeter A, Schickhardt C, Jungkunz M, Bergbold S, Mehlis K, Winkler EC. Patients' willingness to provide their clinical data for research purposes and acceptance of different consent models. Findings from a representative survey of cancer patients in Germany. NCT - Section of Translational Medical Ethics. 2022. URL: https://heidata.uni-heidelberg.de/dataset.xhtml?persistentId=doi:10.11588/data/PJCNTN [accessed 2022-08-10]

Abbreviations

EU: European Union

Edited by R Kukafka; submitted 17.03.22; peer-reviewed by B Leach, J van Delden, K Fultz Hollis; comments to author 14.05.22; revised version received 27.06.22; accepted 11.07.22; published 25.08.22

Please cite as:
Köngeter A, Schickhardt C, Jungkunz M, Bergbold S, Mehlis K, Winkler EC
Patients' Willingness to Provide Their Clinical Data for Research Purposes and Acceptance of Different Consent Models: Findings From a Representative Survey of Patients With Cancer
J Med Internet Res 2022;24(8):e37665
URL: https://www.jmir.org/2022/8/e37665
doi: 10.2196/37665
PMID:

Original Paper

Evaluation of Medical Information on Male Sexual Dysfunction on Baidu Encyclopedia and Wikipedia: Comparative Study

Ming Ma[1], MD; Saifu Yin[1], MA; Mengli Zhu[2], MA; Yu Fan[1], MD; Xi Wen[3], BA; Tao Lin[1], MD; Turun Song[1], MD

[1]Department of Urology and Organ Transplantation Center, West China Hospital, Sichuan University, Chengdu, Sichuan Province, China
[2]Core Facilities, West China Hospital, Sichuan University, Chengdu, Sichuan Province, China
[3]Department of Urology, West China Hospital, Sichuan University, Chengdu, Sichuan Province, China

Corresponding Author:
Turun Song, MD
Department of Urology and Organ Transplantation Center
West China Hospital
Sichuan University
#37 Guoxue Alley
Chengdu, Sichuan Province, 610041
China
Phone: 86 028 854 229 25
Email: songturun1986@scu.edu.cn

Abstract

Background: Sexual dysfunction is a private set of disorders that may cause stigma for patients when discussing their private problems with doctors. They might also feel reluctant to initiate a face-to-face consultation. Internet searches are gradually becoming the first choice for people with sexual dysfunction to obtain health information. Globally, Wikipedia is the most popular and consulted validated encyclopedia website in the English-speaking world. Baidu Encyclopedia is becoming the dominant source in Chinese-speaking regions; however, the objectivity and readability of the content are yet to be evaluated.

Objective: Hence, we aimed to evaluate the reliability, readability, and objectivity of male sexual dysfunction content on Wikipedia and Baidu Encyclopedia.

Methods: The Chinese Baidu Encyclopedia and English Wikipedia were investigated. All possible synonymous and derivative keywords for the most common male sexual dysfunction, erectile dysfunction, premature ejaculation, and their most common complication, chronic prostatitis/chronic pelvic pain syndrome, were screened. Two doctors evaluated the articles on Chinese Baidu Encyclopedia and English Wikipedia. The Journal of the American Medical Association (JAMA) scoring system, DISCERN instrument, and Global Quality Score (GQS) were used to assess the quality of disease-related articles.

Results: The total DISCERN scores (P=.002) and JAMA scores (P=.001) for Wikipedia were significantly higher than those of Baidu Encyclopedia; there was no statistical difference between the GQS scores (P=.31) for these websites. Specifically, the DISCERN Section 1 score (P<.001) for Wikipedia was significantly higher than that of Baidu Encyclopedia, while the differences between the DISCERN Section 2 and 3 scores (P=.14 and P=.17, respectively) were minor. Furthermore, Wikipedia had a higher proportion of high total DISCERN scores (P<.001) and DISCERN Section 1 scores (P<.001) than Baidu Encyclopedia. Baidu Encyclopedia and Wikipedia both had low DISCERN Section 2 and 3 scores (P=.49 and P=.99, respectively), and most of these scores were low quality.

Conclusions: Wikipedia provides more reliable, higher quality, and more objective information than Baidu Encyclopedia. Yet, there are opportunities for both platforms to vastly improve their content quality. Moreover, both sites had similar poor quality content on treatment options. Joint efforts of physicians, physician associations, medical institutions, and internet platforms are needed to provide reliable, readable, and objective knowledge about diseases.

(J Med Internet Res 2022;24(8):e37339) doi: 10.2196/37339

KEYWORDS

sexual dysfunction; digital health; Baidu Encyclopedia; Wikipedia; internet; health information; DISCERN instrument

Introduction

Knowledge regarding health and well-being is cobbled together from health care professionals, family, friends, books, newspapers, magazines, educational pamphlets, radio, television, and pharmaceutical advertisements [1]. However, we are increasingly heading online for answers rather than pursuing information through these other avenues [2]. Approximately 6% of all internet searches in the United States are health-related [3], and it is believed that internet searches have become people's first choice of method to seek information regarding health issues [4]. In addition, the population of netizens in mainland China reached 1011 million in 2021, and the number of online medical users in China had reached 239 million by June 2021, accounting for 23.7% of total internet users [5]. Information quality, emotional support, and source credibility have significant and positive impact on the likelihood of health care information adoption, and among these factors, information quality has the biggest impact on patients' adoption decisions [6]. Given the large amount of inaccurate information online, users are very easily misinformed [1]. Previous studies showed that the quality of online health information is problematic [7,8]. Thus, the assessment of source reputability and the veracity of information is a crucial and urgent task.

As the most common male sexual dysfunctions, erectile dysfunction (ED; the persistent inability to attain and maintain an erection sufficient to permit satisfactory sexual performance) and premature ejaculation (PE; poorly controlled and rapid ejaculation) greatly affect the quality of life of patients [9,10]. Furthermore, sexual dysfunction is closely associated with chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS; urologic pain or discomfort in the pelvic region associated with lower urinary tract symptoms) and is the most common complication [11,12]. The prevalence of CP/CPPS in men is about 8.2%, and men with CP/CPPS are more prone to ED and PE than the general population [13]. A previous study found that nearly half of patients with a self-reported diagnosis of CP/CPPS reported mild to severe ED [14]. A meta-analysis of 24 studies suggested that the overall prevalence of sexual dysfunction in patients with CP/CPPS was 0.62 [15]. In particular, our previous study found that "prostate" and "prostatitis" were the most queried terms by Chinese users with PE [16], which highlighted the stigma and preferences of these patients [17]. In addition, the complex and unclear etiology of CP/CPPS and sexual dysfunction not only challenges clinicians in the choice of treatment but also seriously affects the quality of life of patients. Previously, public interest and the change over time in the search volume for sexual dysfunctions and lower urinary tract symptoms were analyzed [16,18,19]. People tended to consult Dr. Internet in a combined manner on these issues for treatment decision-making. Therefore, the issue of sexual dysfunction is commonly investigated with CP/CPPS.

Wikipedia, the most popular and consulted encyclopedia website in English, is a web-based encyclopedia that provides valuable web-based health information [20]. Previous studies have shown that Wikipedia is a reasonably reliable medical resource and it was ranked higher on search engines than other general websites [21,22]. Unfortunately, on May 19, 2015, "Chinese Wikipedia" announced that mainland Chinese servers would be shut down because of violation of mainland China's laws due to the attack and destruction of the internet. As the equivalent Wikipedia for Chinese internet users, the Baidu platform and its Encyclopedia service is the most popular and frequently consulted encyclopedia site in mainland China [23,24]. In mainland China, with 766 million users actively using the Baidu search service, its usage in relation to health inquiries and symptom confirmation accounts for 66.83% of use, and health and medical topics ranked first among science topics [24]. Our previous research on the Baidu search index showed that the search demands by its users for sexual dysfunction and lower urinary tract symptoms are huge. However, users often get irrelevant online medical information, and there is little evaluation of the quality of Baidu-related content [16,19]. The purpose of this paper was to assess the reliability, readability, and objectivity of Wikipedia and Baidu Encyclopedia content on ED, PE, and CPPS/CP for the advancement of internet medicine.

Methods

Data Sources

The contents analyzed in this study are available on Chinese Baidu Encyclopedia and English Wikipedia. The Chinese Baidu Encyclopedia and English Wikipedia were investigated for articles on ICD-10 version 2016 codes. All possible synonymous and derivative keywords for each term were screened. Two doctors evaluated the articles on Chinese Baidu Encyclopedia and English Wikipedia. Any disagreement was reviewed by and arbitrated by a third reviewer who was an expert on sexual dysfunction. All authors have many years of experience in andrology and urology and are competent in the diagnosis and treatment of male sexual dysfunction and urinary disorders. These reviewers have professional knowledge of male sexual dysfunction and urinary disorders and can make professional evaluations.

Assessment of the Quality of the Research Articles

The Journal of the American Medical Association (JAMA) scoring system [25], DISCERN instrument [26], and Global Quality Score (GQS) [27] were used to assess the quality of disease-related articles. The contents of these scoring tables are shown in Multimedia Appendix 1. The JAMA scoring system is a well-known tool for evaluating the quality of information obtained from health-related websites. It includes 4 evaluation dimensions: author, attribution, disclosure, and currency. If it meets the requirements of each dimension, it will get 1 point, and the deimension with the highest quality will get 4 points. The DISCERN instrument has been developed to judge the quality of written health information [26]. To more comprehensively determine the quality of information in the article, the DISCERN tool consists of 15 questions plus an overall quality rating, and each is scored on a scale from 1 to 5. The first section of the DISCERN instrument is commonly used to evaluate the quality of published information, and the second section focuses on the quality of treatment choices offered to patients. The total score can range from 16 to 80, where a score of 63 to 80 suggests excellent quality, 51 to 62 indicates good quality, 39 to 50 indicates fair quality, and 16 to 38 indicates poor quality [26]. Experienced health information users and providers can use the DISCERN instrument to distinguish between high-quality and low-quality publications, so as to promote the generation of high-quality, evidence-based patient information. The GQS is a 5-point Likert scale that can subjectively rate the overall quality of each reviewed website. In addition to evaluating the overall quality of the website, GQS also considers the flow and ease of use of each website [28].

Statistical Analysis

All databases were constructed with Excel 2019 (Microsoft Corporation, Redmond, WA). The Shapiro-Wilk test was used to test the normality of the data. Descriptive analyses are reported as means and SDs for normally distributed variables. Medians and IQRs are reported for non-normally distributed variables. To ensure the quality of these scores, the intraclass correlation coefficient (ICC) was used to evaluate interobserver reliability. ICC values range from 0 (untrusted) to 1 (fully trusted), and any concordance values less than 0.75 were discussed by the research team to clarify the discrepancy. For nonparametric tests, the Mann-Whitney *U* test was conducted to test the significance of different ranks by using SPSS, version 22.0 (IBM Corp, Armonk, NY). The Fisher exact test was used to test the difference in the frequency distribution of DISCERN scores. We used Prism 8 for macOS, version 8.4.0 (455; GraphPad Software Inc, San Diego, CA) to conduct statistical analyses and create figures. For the statistical analysis, $P<.05$ was considered significant.

Results

Content Characteristics

We searched for “erectile dysfunction,” “premature ejaculation,” “chronic prostatitis/chronic pelvic pain syndrome,” and similar keywords on English Wikipedia and Chinese Baidu Encyclopedia. The search results are shown in Table 1. Wikipedia has only 1 entry for a disease, corresponding to a specific article. In Baidu Encyclopedia, a disease may have multiple entries and multiple articles. The information sources of these articles are different, and the number of views varies greatly. In Wikipedia, an article about a disease is constantly supplemented by different registered individuals. However, Baidu Encyclopedia's content providers are official organizations or unregistered individuals. Moreover, some of the recently updated articles in Baidu Encyclopedia show that the information is more often provided by organizations or institutions and is certified by experts. In addition, both Baidu and Wikipedia provide links to external information, including videos, articles, and images, while some links are unrelated advertisements. The latter especially appear in Baidu Encyclopedia. Furthermore, Wikipedia provides its own features for assessing the quality of articles, and all Wikipedia articles included in this study were rated as grade C.

Table 1. Characteristics of the search results from 2 online platforms.

Themes	Wikipedia	Baidu Encyclopedia	P value[a]
Available entries, n			
CPPS/CP[b]	2	3	N/A[c]
ED[d]	1	2	
PE[e]	1	3	
Real-time updates, n			
Yes	4	8	.99
No	0	0	
External links, n			
Yes	4	6	.52
No	0	2	
Advertisement, n			
Yes	0	4	.21
No	4	4	
Author type, n			
Organization	4	5	.49
Individuals	0	3	
Page views (x1000), median (IQR)	1673.2 (240.0-3878.9)	4119.7 (775.3-22029.8)	.37
Number of references, median (IQR)	53.5 (19.3-84.0)	0 (0.0-0.8)	.002

[a] A Mann-Whitney *U* test was conducted to test the significance of different ranks.

[b] CP/CPPS: chronic prostatitis/chronic pelvic pain syndrome.

[c] N/A: not applicable.

[d] ED: erectile dysfunction.

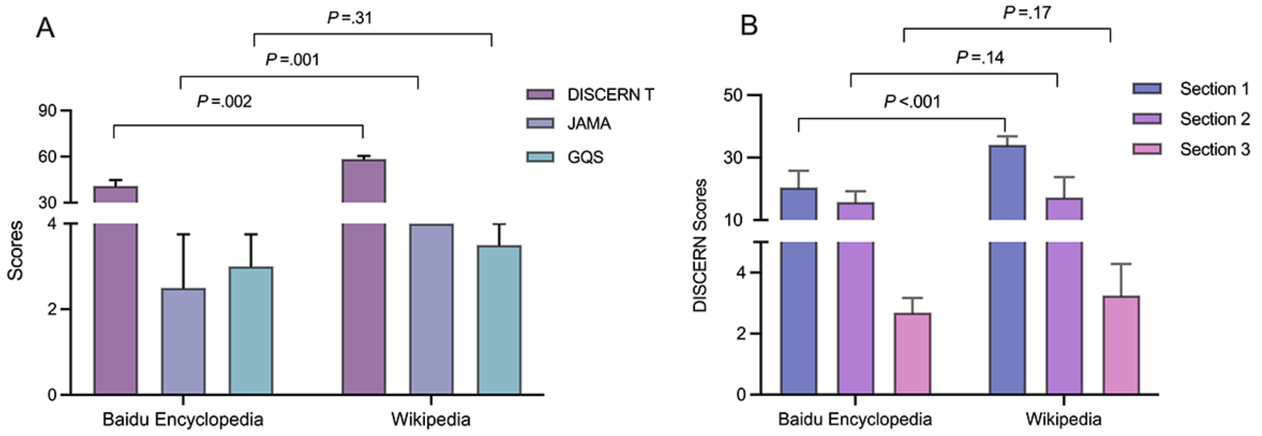
[e] PE, premature ejaculation.

Overall Scores for Baidu Encyclopedia and Wikipedia

A 2-way mixed/random effects model was used to analyze the consistency of the ratings by the 2 independent reviewers. The ICC results showed good consistency between the 2 reviewers for the GQS scores (ICC=0.87), JAMA scores (ICC=0.91), and DISCERN scores (ICC=0.82).

Comprehensively, the scores for Wikipedia were higher than those for Baidu Encyclopedia (Figure 1A). The contents in Wikipedia were significantly higher rated by the DISCERN tool and JAMA tool than those in Baidu Encyclopedia, suggesting that Wikipedia provides higher quality information. Although there was no statistical difference between the GQS scores for these websites, a numerically higher score on Wikipedia indicates that Wikipedia may provide better reading fluency and ease of use. In order to distinguish the differences between the 2 websites in more detail, we compared the DISCERN section scores for Baidu Encyclopedia and Wikipedia (Figure 1B). The DISCERN Section 1 score for Wikipedia was significantly higher than that for Baidu Encyclopedia, suggesting that Wikipedia provides more reliable and more objective information. The DISCERN Section 2 evaluates “How good is the quality of information regarding treatment choices?” There was no statistical difference between the DISCERN Section 2 scores for these websites, suggesting that they may have a similar impact on patients’ choice of treatment options. Section 3 is the overall rating of the publication, and the lack of statistical difference revealed that the overall quality of the publication as a source of information about treatment choices was similar for these websites.

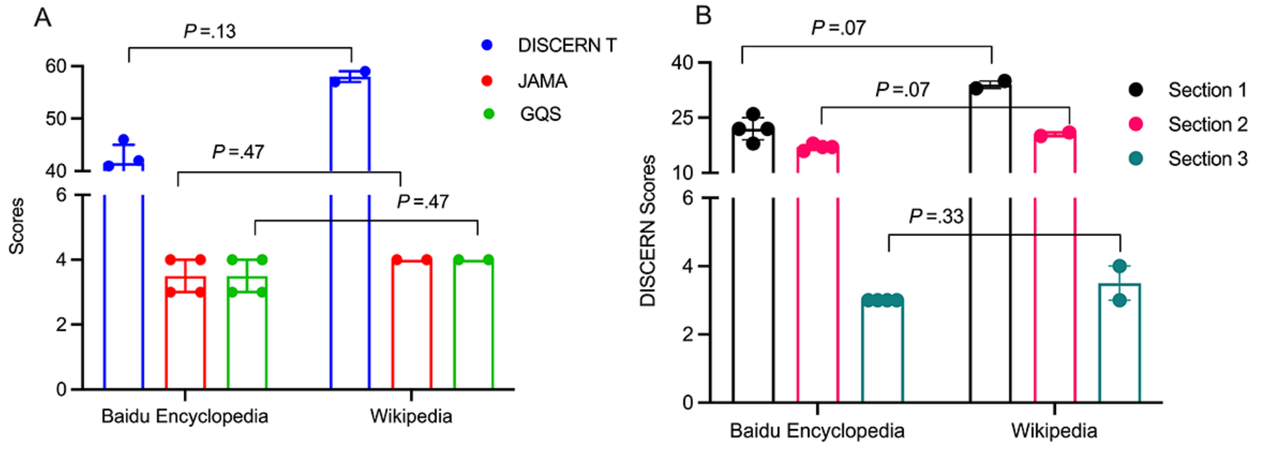
Figure 1. Overall comparison between Baidu Encyclopedia and Wikipedia: (A) median and IQR for DISCERN total scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS) scores; (B) median and IQR for the 3 DISCERN sections.



Overall Quality Comparison Between Wikipedia and Baidu Encyclopedia for the Theme of ED

ED is one of the most common male sexual dysfunctions. By comparing the content scores for ED articles on Baidu Encyclopedia and Wikipedia, Wikipedia appeared to have numerically higher total DISCERN scores, JAMA scores, and GQS scores, but there were no statistically significant differences (Figure 2A). Furthermore, the 3 DISCERN section scores for Baidu and Wikipedia were also compared separately (Figure 2B). Wikipedia appeared to have numerically higher DISCERN Section 1 and 2 scores. In addition, they had similar DISCERN Section 3 scores. These results suggest that there is no statistically significant difference between Wikipedia and Baidu Encyclopedia scores for ED content.

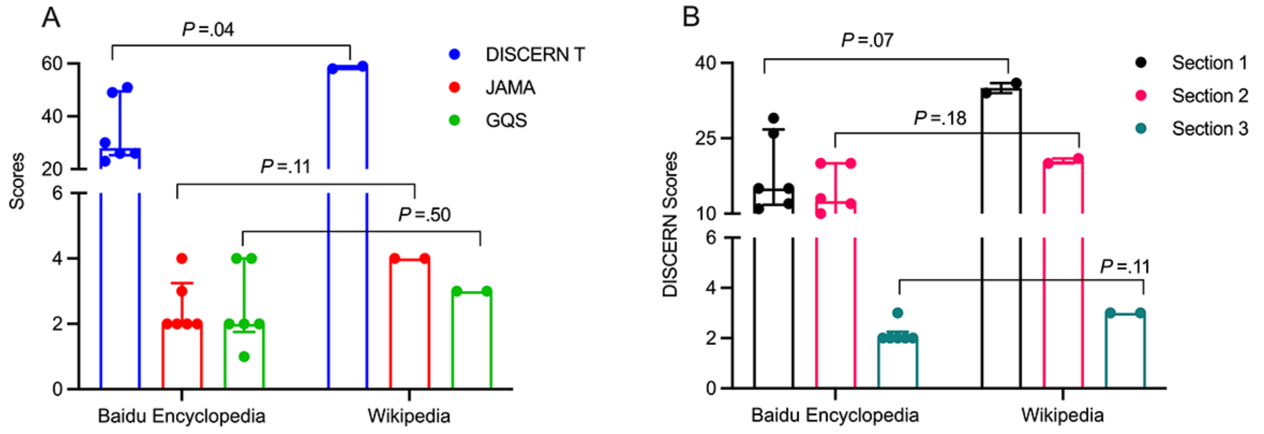
Figure 2. Comparison of erectile dysfunction (ED) scores between Baidu Encyclopedia and Wikipedia: (A) median and IQR for total DISCERN scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS); (B) median and IQR for the 3 DISCERN sections.



Overall Quality Comparison Between Wikipedia and Baidu Encyclopedia for the Theme of PE

A comparison of the scores for PE, the other most common sexual dysfunction disorder, showed that Wikipedia had a significantly higher total DISCERN score than Baidu Encyclopedia (Figure 3A). Although Wikipedia seemed to have higher JAMA and GQS scores than Baidu Encyclopedia (Figure 3A), this difference was not statistically significant, and all DISCERN section scores showed a similar trend (Figure 3B), which may be related to the great intragroup variability of Baidu Encyclopedia.

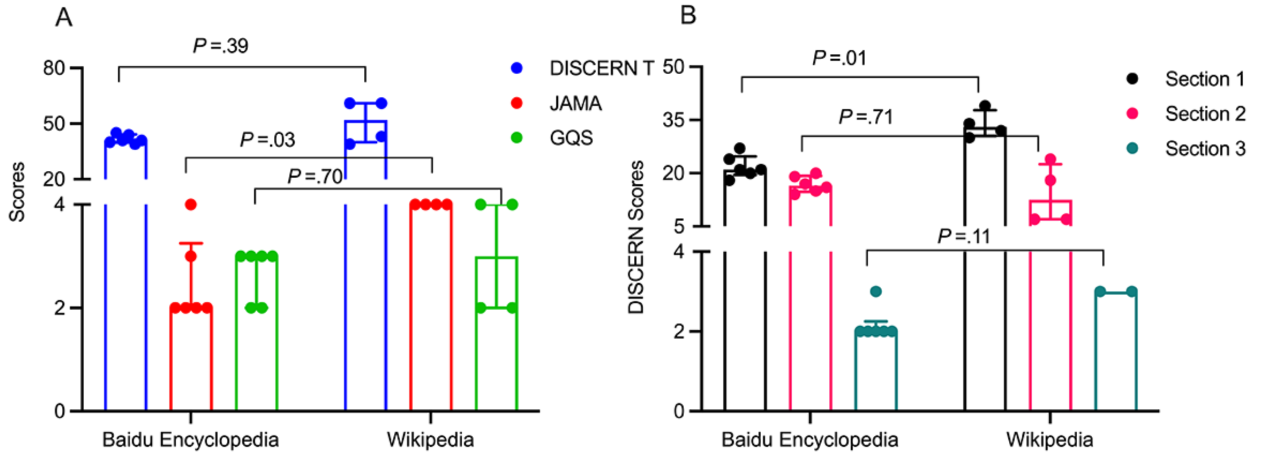
Figure 3. Comparison of premature ejaculation (PE) scores between Baidu Encyclopedia and Wikipedia: (A) median and IQR for total DISCERN scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS); (B) median and IQR for the 3 DISCERN sections.



Overall Quality Comparison Between Wikipedia and Baidu Encyclopedia for the Theme of CP/CPPS

CP/CPPS, as one of the most common concomitant diseases of sexual dysfunction, seriously affects the quality of life of male patients. By comparing the overall scores for Wikipedia and Baidu encyclopedia on CP/CPPS, we found that the scores of Baidu Encyclopedia were mostly fair quality, while the scores of Wikipedia ranged from fair quality to good quality (Figure 4A). Meanwhile, Wikipedia showed statistically higher JAMA scores, but there were no statistical differences between total DISCERN scores and GQS scores (Figure 4A). Furthermore, the DISCERN Section 1 score for Wikipedia was statistically significantly higher than that of Baidu Encyclopedia, while the DISCERN Section 2 and 3 scores for both sites were not significantly different from each other (Figure 4B).

Figure 4. Comparison of chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS) scores between Baidu Encyclopedia and Wikipedia: (A) median and IQR for total DISCERN scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS); (B) median and IQR for the 3 DISCERN sections.



Distribution of the DISCERN Scores

After comparing the overall quality of the information for different diseases on Baidu encyclopedia and Wikipedia, the overall scores for Wikipedia seemed to be higher than those of Baidu encyclopedia, but some scores only showed numerical differences without statistical significance. Nevertheless, the differences in the distribution of scores that had numerical differences were seemingly obvious. Therefore, we performed further statistical analyses of the score distributions for Wikipedia and Baidu Encyclopedia. As aforementioned, according to the DISCERN standard, a total DISCERN score <50 (near 60%) is fair or poor quality, while a score >50 is good or excellent quality [26]. Based on this rule, we took a score of 3 for each question as the cutoff value; that is, a score higher than 3 points was defined as good quality.

The score distributions for each disease are shown in Table 2. Wikipedia had a higher proportion of total DISCERN and Section 1 scores distributed above 3 points, whether compared with the overall score or the score for each disease, and was significantly better than Baidu Encyclopedia. However, Baidu Encyclopedia and Wikipedia had low Section 2 and 3 scores, and most of these scores were ≤3, which are defined as low quality.

Table 2. Distribution of the DISCERN scores for each disease and comparisons via the Fisher exact test.

DISCERN	Overall			CP/CPPS[a]			ED[b]			PE[c]		
	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value
Total												
>3	38 (59.4)[d]	25 (19.5)[e]	<.001	17 (53.1)[f]	10 (20.8)[g]	.004	10 (62.5)[h]	7 (21.9)[f]	<.001	11 (68.8)[h]	8 (16.7)[g]	<.001
≤3	26 (40.6)[d]	103 (80.5)[e]		15 (46.9)[f]	38 (79.2)[g]		6 (37.5)[h]	25 (78.1)[f]		5 (31.2)[h]	40 (83.3)[g]	
Section 1												
>3	31 (96.9)[f]	19 (29.7)[d]	<.001	15 (93.8)[h]	7 (29.2)[i]	<.001	8 (100)[j]	6 (37.5)[h]	.006	8 (100)[j]	6 (25.0)[i]	<.001
≤3	1 (3.1)[f]	45 (70.3)[d]		1 (6.2)[h]	17 (70.8)[i]		0 (0)[j]	10 (62.5)[h]		0 (0)[j]	18 (75.0)[i]	
Section 2												
>3	5 (17.9)[k]	6 (10.7)[l]	.49	1 (7.1)[m]	3 (14.3)[n]	.64	1 (14.3)[o]	1 (7.1)[m]	.99	3 (42.9)[o]	2 (9.5)[n]	.08
≤3	23 (82.1)[k]	50 (89.3)[l]		13 (92.9)[m]	18 (85.7)[n]		6 (85.7)[o]	13 (92.9)[m]		4 (57.1)[o]	19 (90.5)[n]	
Section 3												
>3	2/4 (50.0)[p]	3 (37.5)[j]	.99	1 (50.0)[q]	0 (0)[r]	.40	1 (100)[s]	0 (0)[q]	.33	0 (0)[s]	0 (0)[r]	.99
≤3	2 (50.0)[p]	5 (62.5)[j]		1 (50.0)[q]	3 (100)[r]		0 (0)[s]	2 (100)[q]		1 (100)[s]	3 (100)[r]	

[a] CP/CPPS: chronic prostatitis/chronic pelvic pain syndrome.

[b] ED: erectile dysfunction.

[c] PE: premature ejaculation.

[d] n=64.

[e] n=128.

[f] n=32.

[g] n=48.

[h] n=16.

[i] n=24.

[j] n=8.

[k] n=28.

[l] n=56.

[m] n=14.

[n] n=21.

[o] n=7.

[p] n=4.

[q] n=2.

[R] n=3.

[S] n=1.

Discussion

Principal Findings

Internet-based information is playing an increasingly important role in the diagnosis and treatment of patients, especially for privacy-sensitive conditions such as sexual dysfunction and related concomitant diseases. Comprehensive and objective information can help patients understand their condition, choose the right time to visit a doctor, and then improve their prognosis. However, incorrect or incomplete information may leave patients vulnerable to misdiagnosis, leading to delays in treatment and considerable health risks [1]. As a consequence, at a time when internet health care is booming, there is an urgent need to evaluate the credibility, readability, and accuracy of online resources. This study evaluated the reliability, readability, and objectivity of Baidu Encyclopedia and Wikipedia in terms of ED, PE, and CP/CPPS content. Overall, the total DISCERN scores and DISCERN Section 1 scores for the content provided by Wikipedia were significantly higher than those of Baidu Encyclopedia. Also, Wikipedia had a higher proportion of total DISCERN and Section 1 scores distributed within the high-quality range than Baidu Encyclopedia. Combined with higher JAMA scores, the results suggest that Wikipedia provided more reliable, higher quality, and more objective information than Baidu Encyclopedia. Baidu Encyclopedia and Wikipedia had low DISCERN Section 2 and 3 scores, and most of these scores were low quality. Similar DISCERN Section 2 and 3 scores for Wikipedia and Baidu Encyclopedia indicated that they had an analogic and mediocre impact on patients' choice of treatment options. Although not statistically different, Wikipedia had numerically higher GQS scores, suggesting that Wikipedia might provide relatively better flow and be easier to use.

By June 2021, the number of online medical users in China was 239.33 million, and the utilization rate of the internet was 23.7%, an increase of 11.4% over December 2020 [29]. In an analysis of internet search trends in China, some scholars found that only 43.74% of the search results for PE were related to PE [16]. In another study on lower urinary tract symptoms, 1.13%-93.92% of the retrieved content was found to be irrelevant to lower urinary tract symptoms [19]. The study also found similar problems in the contents about these diseases in Wikipedia and Baidu Encyclopedia. Wikipedia provides more standardized and unified content, with standard templates for almost every disease, which allows readers to find the information they need quickly and accurately [30]. In contrast, the quality of content provided by Baidu Encyclopedia varies widely, with some recently updated articles providing more comprehensive content than Wikipedia, but the overall trend is a lack of standardization and formality. In Baidu Encyclopedia, the same disease may correspond to multiple entries and corresponding articles, which compare poorly with each other, and different articles may provide users with contradictory information, which can cause great confusion to users. The diversity of the content formats presented by Baidu Encyclopedia is consistent with the great variability of its overall score. The total DISCERN scores and JAMA scores for Wikipedia were significantly higher than those for Baidu Encyclopedia, and the proportion of Wikipedia scores within the high-quality distribution was also higher than those for Baidu Encyclopedia. These results suggest that Wikipedia provides higher quality information than Baidu Encyclopedia. In addition to the lack of a standard content presentation format, the low quality of Baidu Encyclopedia is also related to other features of its website, such as information sources and references. The contents of Baidu Encyclopedia are mostly sourced from official organizations or unregistered individuals, while information on Wikipedia is provided by registered users. The comparison shows that the quality of contents provided by unregistered individuals is always rated as "poor quality." Accurate citation of high-quality references is an important guarantee for the reliability of a paper [31]. The contents provided by these unregistered individual users are almost always without references and extended information. By contrast, the quality of contents provided by registered users or official organizations are almost rated as "good quality," with accurate references. These characteristics of the website are closely related to DISCERN Section 1 scores, and significantly higher DISCERN Section 1 scores for Wikipedia indicate that its publications are more reliable than those of Baidu Encyclopedia. The other 2 main focuses of the quality assessment are "How good is the quality of information regarding treatment choices?" and "the overall quality of the publication as a source of information about treatment choices." Similar scores on DISCERN Sections 2 and 3 for Wikipedia and Baidu Encyclopedia indicated that they had an analogic and mediocre impact on patients' choice of treatment options. Recent updates to Baidu Encyclopedia also show an increasing number of medical professionals involved in reviewing or writing the content, also significantly improving the DISCERN and JAMA scores. This comparison suggests that the inconsistency of disease presentation formats and differences in information sources may account for the lower Baidu scores.

CP/CPPS is characterized by localized pain or discomfort in the abdomen, pelvis, and genitals, usually with lower urinary tract symptoms, psychosocial disorders, and sexual dysfunction [11,12]. The relationship between sexual dysfunctions and CP/CPPS has been studied more extensively [32]. Previous studies have shown a good correlation between the severity of symptom scores between the 2 clinical conditions, CP and PE, and that approximately 49% of male patients with CP have concomitant sexual dysfunction [33]. It addition, "prostate" and "prostatitis" were the most queried terms by Chinese users with PE [16]. The complex and heterogeneous pathophysiology of CP/CPPS makes the management of this troublesome situation very challenging both for clinicians and patients, and approximately 50% of older patients experience recurrence [34]. The UPOINT System classifies CP/CPPS patients into 7 different subgroups based on symptoms: urologic, psychosocial, organ-specific, infectious, neurologic, tenderness (pelvic floor tenderness), and sexual dysfunction; then, it proposes specific treatment plans based on the different subgroups [35]. There is growing evidence that the addition of second-line therapies, such as 5-phosphodiesterase inhibitors, antidepressants and muscle relaxants, according to the UPOINT System approach, can significantly improve patients' CP/CPPS symptoms [36]. These results showed that CP/CPPS and sexual dysfunctions can directly or indirectly increase the economic burden of health care and seriously affect patients' quality of life. Patients with CP/CPPS or sexual dysfunction may feel too embarrassed to discuss their problems with doctors due to the influence of the Chinese culture, and they are likely more willing to look for disease-related information, such as symptoms, diagnosis, treatment methods, prognosis, and hospital rankings, on the internet first. There is no doubt that the information these patients access from the internet affects their perception of their health status, which in turn affects treatment choices and disease prognosis.

By comparing the contents for ED, PE, and CP/CPPS on Baidu Encyclopedia and Wikipedia, we found that the consistency of Wikipedia is better, with almost all content rated as "good quality," while the scores for Baidu Encyclopedia were mostly "fair quality." Take PE-related articles in Baidu Encyclopedia as examples. Both reviewers rated "早发性射精" (early-onset ejaculation) as “poor quality.” After analyzing the content on the web page for “early-onset ejaculation,” we found there was no introduction to “examination, diagnosis, and treatment,” and the content in the article was not objective and scientific. Contrary to the lack of effective information, there are more than 25 irrelevant advertising links and only one reference on this web page. The content on the “早泄” (premature ejaculation) page on Baidu Encyclopedia was rated as “good quality,” and the information was more comprehensive and objective than that for “early-onset ejaculation.” Corresponding to the quality grades for “early-onset ejaculation” and “premature ejaculation,” there was a huge difference in page views (early-onset ejaculation/premature ejaculation: 33,506/25,747,398). The discrepancy may be related to the inconsistent identity of content providers. The irrelevant advertising links or misleading information obtained by users using Baidu Encyclopedia may be related to the fee-based editing service. There are many third-party underground industries that charge fees to write Baidu Encyclopedia entries on their behalf, so as to insert advertisements and achieve the purpose of attracting patients. In order to improve the quality of the health information, Baidu Encyclopedia announced the “rainbow plan” on December 9, 2012, wherein all medical entries could only be edited and revised by certified medical experts [37]. This is consistent with the findings of this study that an increasing number of medical professionals are involved in reviewing or writing content for Baidu Encyclopedia. Consequently, attracting, encouraging, and even recruiting more medical professionals to draft or proofread the content about disease presentation provided on these websites may ensure the content is objective and comprehensive. At the same time, the Baidu Encyclopedia platform should strengthen content regulation and establish a review mechanism to remove interest-related content.

In contrast, Wikipedia has its own content quality evaluation system, such as the “Wiki-Project article quality grading scheme” and the “Wiki-Project priority assessments” [38,39]. In this study, all included Wikipedia articles were rated as grade C, which means “Useful to a casual reader, but would not provide a complete picture for even a moderately detailed study” and “Considerable editing is needed to close gaps in content and solve cleanup problems.” The “Wiki grading” for these Wikipedia articles is similar to the grading by the 3 grading tools applied in this paper. That is, the quality of these Wikipedia articles is almost “good quality” but far from “excellent quality,” and all articles needed further improvement. Despite this fact, the formality and drafting on Wikipedia are better because of the clear attribution and disclosure it provides. As mentioned earlier, there is a lack of uniform standards for writing Baidu Encyclopedia content, many of the information sources are not supported by academic references, and external links are mostly related to advertisements. Hence, though the content on both sites leaves much to be desired, as a source to popularize science, the content on Wikipedia could at least guide interested individuals to the right source of informations, while Baidu Encyclopedia is more likely to provide misleading information.

In the era of rapid internet development, more patients have started to try online consultations [40]. This change in mode of treatment has presented new opportunities and challenges for doctors, medical institutions, physician associations, internet platforms, and patients. In this study, we evaluated the objectivity, reliability, and readability of the content on sexual dysfunction and CP/CPPS on Baidu and Wikipedia and found that the quality of the content provided by both sites was not “excellent quality” and needed to be improved. This study is only a microcosm of the vast amount of information available in internet-based health care. Considering the increasing coverage of the internet, more users will be influenced by internet-based information, and incorrect or incomplete information will have a negative impact on users’ decision-making. Therefore, we believe that, in the era of the internet information explosion, physicians, physician associations, and medical institutions should make full use of their expertise and become more involved in the construction of internet-based health care by providing objective and comprehensive content. Internet platforms, on the other hand, should strengthen the regulation and review of medical-related content and remove false or irrelevant content. Wikipedia already has a relatively complete self-censorship system and self-evaluation system, but Baidu Encyclopedia has almost no achievements in this regard. In China, the country with the world’s largest population, the importance of popular science education for the whole society and the world is self-evident. Baidu Encyclopedia, as the largest platform for online science education in China, still needs to be greatly enhanced to take up the corresponding social responsibility. Through the joint efforts of physicians and the platform, we hope to achieve the goal of providing users with timely access to correct, objective, comprehensive, and valid information when seeking medical advice or searching for health science content on the internet.

Limitations

Some limitations must be addressed in this study. This study only presents the results of medical professionals’ evaluations of health-related science content on the internet, and further research is needed on the specific impact of this information on the audience and readers. Since information on the internet is updated quickly, there may be some bias between the study results and the actual situation, and the data need to be updated in real time to ensure that the findings are true and valid. In addition, the difference in the number of Chinese and English entries indicates the information received by users will be significantly different because of the entries they choose to click. Therefore, our “combined” evaluation cannot fully represent the quality of the information they really receive. Fortunately, with the availability of infodemiology research, academics can combine content analysis and infodemiology search trends to better elucidate the impact of health-related information on the internet on users, society, and the health care industry.

Conclusions

Internet medicine, as a new medical model in the new era, provides strong support for users to understand disease information and choose the timing of treatment in a timely manner. Although it is more formally composited, Wikipedia also provides more reliable, higher quality, and more objective information than Baidu Encyclopedia. They also have a similar impact on patients' choice of treatment options, and the websites are similar in terms of flow and ease of use. To promote the healthy and sustainable development of internet health care, the joint efforts of physicians, physician associations, medical institutions, and internet platforms are needed to provide more reliable, accessible, and comprehensible disease knowledge to the public.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant numbers 81870513, .81600584, 81470980), 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (grant number ZY2016104).

Authors' Contributions

This study was conceptualized by MM and TL. The methodology was designed by MM, YF, and TS. The investigation was carried out by MM and YF. MM and SY carried out the data statistics and interpretation. MM and MZ wrote the original draft. MM, YF, TS, and TL reviewed and edited the draft. Funding was acquired by TL. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Scoring tables for the Journal of the American Medical Association (JAMA) scoring system, DISCERN instrument, and Global Quality Score (GQS).
[PDF File (Adobe PDF File), 137 KB-Multimedia Appendix 1]

References

1. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41(1):433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
2. Johnson S, Park H, Gross C, Yu J. Use of alternative medicine for cancer and its impact on survival. J Natl Cancer Inst 2018 Jan 01;110(1):1. [doi: 10.1093/jnci/djx145] [Medline: 28922780]
3. Asch JM, Asch DA, Klinger EV, Marks J, Sadek N, Merchant RM. Google search histories of patients presenting to an emergency department: an observational study. BMJ Open 2019 Feb 20;9(2):e024791 [FREE Full text] [doi: 10.1136/bmjopen-2018-024791] [Medline: 30787088]
4. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
5. The 48th "Statistical Report on Internet Development in China". China Internet Network Information Center. 2021 Sep 15. URL: https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202109/t20210915_71543.htm [accessed 2022-02-11]
6. Jin J, Yan X, Li Y, Li Y. How users adopt healthcare information: An empirical study of an online Q&A community. Int J Med Inform 2016 Feb;86:91-103. [doi: 10.1016/j.ijmedinf.2015.11.002] [Medline: 26616406]
7. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
8. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
9. Salonia A, Bettocchi C, Boeri L, Capogrosso P, Carvalho J, Cilesiz N, EAU Working Group on Male Sexual and Reproductive Health. European Association of Urology Guidelines on Sexual and Reproductive Health-2021 Update: Male Sexual Dysfunction. Eur Urol 2021 Sep;80(3):333-357. [doi: 10.1016/j.eururo.2021.06.007] [Medline: 34183196]
10. Wei S, Wu C, Yu B, Ma M, Qin F, Yuan J. Advantages and limitations of current premature ejaculation assessment and diagnostic methods: a review. Transl Androl Urol 2020 Apr;9(2):743-757 [FREE Full text] [doi: 10.21037/tau.2019.12.08] [Medline: 32420180]
11. Krieger JN, Nyberg L, Nickel JC. NIH consensus definition and classification of prostatitis. JAMA 1999 Jul 21;282(3):236-237. [doi: 10.1001/jama.282.3.236] [Medline: 10422990]
12. Magistro G, Wagenlehner FM, Grabe M, Weidner W, Stief CG, Nickel JC. Contemporary management of chronic prostatitis/chronic pelvic pain syndrome. Eur Urol 2016 Feb;69(2):286-297. [doi: 10.1016/j.eururo.2015.08.061] [Medline: 26411805]
13. Tran CN, Shoskes DA. Sexual dysfunction in chronic prostatitis/chronic pelvic pain syndrome. World J Urol 2013 Aug;31(4):741-746. [doi: 10.1007/s00345-013-1076-5] [Medline: 23579441]
14. Matta R, Doiron C, Leveridge MJ. The dramatic increase in social media in urology. J Urol 2014 Aug;192(2):494-498. [doi: 10.1016/j.juro.2014.02.043] [Medline: 24576656]

15. Li H, Kang D. Prevalence of sexual dysfunction in men with chronic prostatitis/chronic pelvic pain syndrome: a meta-analysis. World J Urol 2016 Jul 6;34(7):1009-1017 [FREE Full text] [doi: 10.1007/s00345-015-1720-3] [Medline: 26546073]
16. Wei S, Ma M, Wen X, Wu C, Zhu G, Zhou X. Online public attention toward premature ejaculation in mainland China: infodemiology study using the Baidu Index. J Med Internet Res 2021 Aug 26;23(8):e30271 [FREE Full text] [doi: 10.2196/30271] [Medline: 34435970]
17. Russo GI, di Mauro M, Cocci A, Cacciamani G, Cimino S, Serefoglu EC, EAU-YAU Men's Health Working Group. Consulting "Dr Google" for sexual dysfunction: a contemporary worldwide trend analysis. Int J Impot Res 2020 Jul 07;32(4):455-461. [doi: 10.1038/s41443-019-0203-2] [Medline: 31591474]
18. Hui J, He S, Liu R, Zeng Q, Zhang H, Wei A. Trends in erectile dysfunction research from 2008 to 2018: a bibliometric analysis. Int J Impot Res 2020 Jul;32(4):409-419 [FREE Full text] [doi: 10.1038/s41443-019-0161-8] [Medline: 31235897]
19. Wei S, Ma M, Wu C, Yu B, Jiang L, Wen X, et al. Using search trends to analyze web-based interest in lower urinary tract symptoms-related inquiries, diagnoses, and treatments in mainland China: infodemiology study of Baidu Index data. J Med Internet Res 2021 Jul 06;23(7):e27029 [FREE Full text] [doi: 10.2196/27029] [Medline: 34255683]
20. Weiner SS, Horbacewicz J, Rasberry L, Bensinger-Brody Y. Improving the quality of consumer health information on Wikipedia: case series. J Med Internet Res 2019 Mar 18;21(3):e12450 [FREE Full text] [doi: 10.2196/12450] [Medline: 30882357]
21. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
22. Thomas G, Eng L, de Wolff JF, Grover S. An evaluation of Wikipedia as a resource for patient education in nephrology. Semin Dial 2013;26(2):159-163. [doi: 10.1111/sdi.12059] [Medline: 23432369]
23. Sun F, Yang F, Zheng S. Evaluation of the liver disease information in Baidu Encyclopedia and Wikipedia: longitudinal study. J Med Internet Res 2021 Jan 18;23(1):e17680 [FREE Full text] [doi: 10.2196/17680] [Medline: 33459597]
24. Report on Chinese Netizens' Search Behavior for Popular Science Demand (Q1 2019). China Association of Science and Technology. 2019 Apr 26. URL: https://www.cast.org.cn/art/2019/4/26/art_1281_94546.html [accessed 2022-02-11]
25. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [doi: 10.1001/jama.1997.03540390074039] [Medline: 9103351]
26. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb 01;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
27. Langille M, Bernard A, Rodgers C, Hughes S, Leddin D, van Zanten SV. Systematic review of the quality of patient information on the internet regarding inflammatory bowel disease treatments. Clin Gastroenterol Hepatol 2010 Apr;8(4):322-328. [doi: 10.1016/j.cgh.2009.12.024] [Medline: 20060070]
28. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
29. Yacob M, Lotfi S, Tang S, Jetty P. Wikipedia in vascular surgery medical education: comparative study. JMIR Med Educ 2020 Jun 19;6(1):e18076 [FREE Full text] [doi: 10.2196/18076] [Medline: 32417754]
30. Good BM, Clarke EL, Loguercio S, Su AI. Building a biomedical semantic network in Wikipedia with Semantic Wiki Links. Database (Oxford) 2012 Mar 20;2012:bar060-bar060 [FREE Full text] [doi: 10.1093/database/bar060] [Medline: 22434829]
31. Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. J Med Internet Res 2015 Mar 04;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]
32. McMahon CG, Jannini EA, Serefoglu EC, Hellstrom WJG. The pathophysiology of acquired premature ejaculation. Transl Androl Urol 2016 Aug;5(4):434-449 [FREE Full text] [doi: 10.21037/tau.2016.07.06] [Medline: 27652216]
33. Liang C, Zhang X, Hao Z, Shi H, Wang K. Prevalence of sexual dysfunction in Chinese men with chronic prostatitis. BJU Int 2004 Mar;93(4):568-570. [doi: 10.1111/j.1464-410x.2003.04662.x] [Medline: 15008731]
34. Clemens JQ, Meenan RT, O'Keeffe Rosetti MC, Kimes T, Calhoun EA. Prevalence of and risk factors for prostatitis: population based assessment using physician assigned diagnoses. J Urol 2007 Oct;178(4 Pt 1):1333-1337. [doi: 10.1016/j.juro.2007.05.140] [Medline: 17706722]
35. Magri V, Boltri M, Cai T, Colombo R, Cuzzocrea S, De Visschere P, et al. Multidisciplinary approach to prostatitis. Arch Ital Urol Androl 2019 Jan 18;90(4):227-248 [FREE Full text] [doi: 10.4081/aiua.2018.4.227] [Medline: 30655633]
36. Magri V, Marras E, Restelli A, Wagenlehner F, Perletti G. Multimodal therapy for category III chronic prostatitis/chronic pelvic pain syndrome in UPOINTS phenotyped patients. Exp Ther Med 2015 Mar;9(3):658-666 [FREE Full text] [doi: 10.3892/etm.2014.2152] [Medline: 25667610]
37. Baidu: Medical entries will not be edited by ordinary people. BJNEWS. 2012 Dec 10. URL: https://www.bjnews.com.cn/finance/2012/12/10/238299.html [accessed 2022-02-11]
38. WikiProject article quality grading scheme. Wikipedia. URL: https://en.wikipedia.org/wiki/Wikipedia:Content_assessment [accessed 2022-02-11]

39. WikiProject priority assessments. Wikipedia. URL: https://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Release_Version_Criteria [accessed 2022-02-11]
40. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec 17;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]

Abbreviations

CP/CPPS: chronic prostatitis/chronic pelvic pain syndrome
ED: erectile dysfunction
GQS: Global Quality Score
ICC: intraclass correlation coefficient
JAMA: Journal of the American Medical Association
PE: premature ejaculation

Edited by JMIRPE Office; submitted 16.02.22; peer-reviewed by A Montilha, S Wei, B Yu, X Zhou; comments to author 03.04.22; revised version received 17.05.22; accepted 05.07.22; published 09.08.22

Please cite as:
Ma M, Yin S, Zhu M, Fan Y, Wen X, Lin T, Song T
Evaluation of Medical Information on Male Sexual Dysfunction on Baidu Encyclopedia and Wikipedia: Comparative Study
J Med Internet Res 2022;24(8):e37339
URL: https://www.jmir.org/2022/8/e37339
doi: 10.2196/37339
PMID:

Original Paper

The Technology Acceptance of Video Consultations for Type 2 Diabetes Care in General Practice: Cross-sectional Survey of Danish General Practitioners

Daniel Cæsar Torp, BMedSci; Annelli Sandbæk, MD, PhD; Thim Prætorius, PhD

Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark

Corresponding Author:
Thim Prætorius, PhD
Steno Diabetes Center Aarhus
Aarhus University Hospital
Palle Juul-Jensens Boulevard 11
Aarhus, 8200
Denmark
Phone: 45 20539002
Email: thipra@rm.dk

Abstract

Background: During the COVID-19 pandemic, video consultations became a common method of delivering care in general practice. To date, research has mostly studied acute or subacute care, thereby leaving a knowledge gap regarding the potential of using video consultations to manage chronic diseases.

Objective: This study aimed to examine general practitioners' technology acceptance of video consultations for the purpose of managing type 2 diabetes in general practice.

Methods: A web-based survey based on the technology acceptance model measuring 4 dimensions—perceived usefulness, perceived ease of use, attitude, and behavioral intention to use—was sent to all general practices (N=1678) in Denmark to elicit user perspectives. The data were analyzed using structural equation modeling.

Results: The survey sample comprised 425 general practitioners who were representative of the population. Structural equation modeling showed that 4 of the 5 hypotheses in the final research model were statistically significant ($P<.001$). Perceived ease of use had a positive influence on perceived usefulness and attitude. Attitude was positively influenced by perceived usefulness. Attitude had a positive influence on behavioral intention to use, although perceived usefulness did not. Goodness-of-fit indices showed acceptable fits for the structural equation modeling estimation.

Conclusions: Perceived usefulness was the primary driver of general practitioners' positive attitude toward video consultations for type 2 diabetes care. The study suggests that to improve attitude and technology use, decision-makers should focus on improving usefulness, that is, how it can improve treatment and make it more effective and easier.

(J Med Internet Res 2022;24(8):e37223) doi: 10.2196/37223

KEYWORDS

video consultations; telemedicine; diabetes; chronic diseases; general practice; technology acceptance; technology acceptance model

Introduction

Background

Technological change and the use of new technologies in health care are driven by objectives to increase access to health care, reduce care costs, coordinate health care, and facilitate chronic disease prevention and management [1]. The COVID-19 pandemic, caused by SARS-CoV-2 infection, has spurred health care systems to rapidly change from delivering in-person care to using different types of web-based care [2-4] such as video consultations [5]. Within the primary care sector, the uptake of video consultations has increased [6], and general practitioners' use of the technology has internationally moved from being used in pilot projects to wider-scale use [7-9]. The care potential of using video consultations in general practice is considered high [10,11], and this technology holds the potential to disrupt how health care is delivered in the primary care sector [12].

The recent uptake of video consultations in general practice is intriguing as the use of new health care technology and its implementation typically takes years [5,13]. This is because digital-first approaches to primary care could increase general practice workload [14] or threaten professional autonomy [15]. Similar to the hospital sector [16,17], knowledge about the impact of video consultations on general practice is in its infancy, and the literature is particularly short on quantitative studies [18]. The nascent literature finds that offering video consultations constitutes a significant change in how health care professionals deliver and patients receive care [19]. Research into factors that influence the implementation of video consultations in routine practice finds that, for instance, training is an important facilitator [20], and hesitance to change is an equally important barrier [21]. Research suggests that general practitioner characteristics (eg, age and sex) do not influence use, although working in larger practices makes it more likely [22,23]. Interaction and communication between patients and general practitioners during video consultations are usually effective [24,25]. However, patients and practitioners report mixed user experiences but with the important point that user ratings depend on the context in which video consultations are used [26-31]. Younger patients were found to be more likely to request or be offered a web-based visit [32].

However, research has not systematically elicited general practitioners' attitudes toward video consultations or their perceptions of the ease of use or usefulness in general practice. This research gap is unfortunate as it is well established in IT literature that attitude and perception influence physicians' use of other types of health care technology such as electronic patient records or telemedicine [33-35]. The technology acceptance model (TAM) has proven to be a robust model through rigorous empirical testing within and beyond health care [36,37]. TAM is capable of studying user attitudes and perceptions and has good predictive power of health technology use [38]. Central to the original TAM [39] and later extensions [40] is that the behavioral intention (BI) to use technology is influenced by users' ratings of perceived usefulness (PU), perceived ease of use (PEOU), and attitude toward the technology. Importantly, BI to use predicts actual user behavior [41,42].

Using the insight that chronic disease prevention and management are key drivers of technological change, this paper studies the potential of using video consultations in general practice to manage type 2 diabetes for 3 reasons. First, type 2 diabetes is a chronic disease for which video consultation appears promising in general practice [43-45]. Second, previous research on the use of video consultations in general practice has mostly studied acute or subacute or out-of-hours care and, to a much lesser extent, the management of chronic care taking place during regular hours [17,25,31]. Third, it is important to find care models capable of delivering high-quality and efficient type 2 diabetes care in general practice [46,47] as the disease prevalence is increasing [48] and people living with type 2 diabetes are at higher risk of developing complications [49].

The aim of this paper is to use TAM to study general practitioners' technology acceptance of video consultations to manage type 2 diabetes in general practice. The hypotheses were that higher levels of attitude, PU, and PEOU positively affect general practitioners' BI to use video consultations to manage type 2 diabetes. Bringing to bear TAM on video consultations in general practice allows exploring the potential of using the technology for a type of chronic care where health care systems need to find new ways of increasing health care access and cutting care costs.

Research Model and Hypotheses

The research model (Figure 1) builds on TAM [39] and posits that general practitioners' perception of the degree to which video consultations used to manage type 2 diabetes are easy to use affects both perceptions of usefulness and attitudes toward using the technology. General practitioners' attitudes are also influenced by their perception of how useful the technology is. Ultimately, general practitioners' intention to use video consultations to manage type 2 diabetes can be explained by their attitude toward the technology and PU. The following develops 5 hypotheses by combining research insights on TAM, general practitioners, and the primary health care domain.

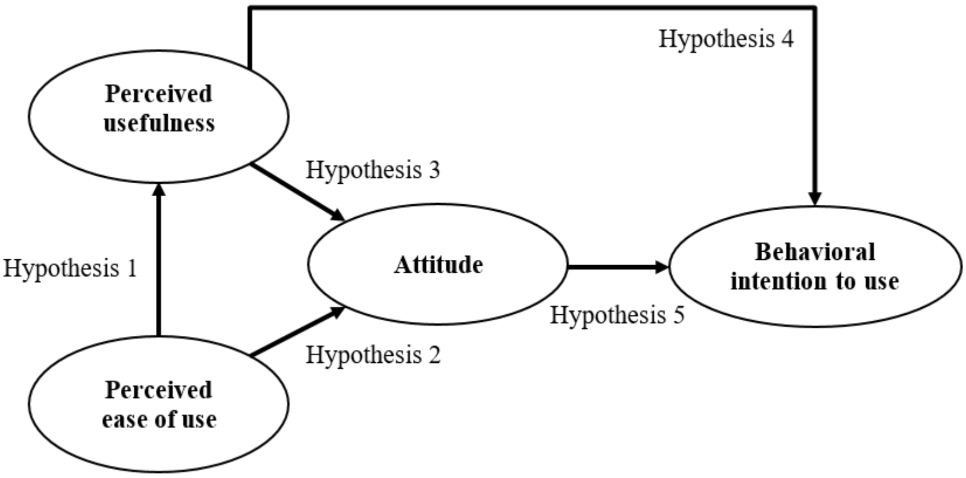
PEOU influences BI to use indirectly through both attitude and PU. A high PEOU represents the belief that using the technology will require little to no effort [39]. PU concerns the extent to which a user believes that the technology can improve or make their work more effective and easier and how it will be advantageous over the current practice. The relationship between PEOU and PU is expected to be positive as health care studies find that a higher level of PEOU leads to higher ratings of P [50-52]. Moreover, studies have shown that when a technology is perceived as easy to use, the attitude toward the technology is more positive [40,52]. The attitudinal component of the model measures an individual's affective response to adopting a new technology. Attitude concerns the extent to which a user finds that using the technology is a good idea, beneficial, or unpleasant for the way they work [39]. PU is considered particularly important in general practice [53,54], and research using TAM finds that physicians' PU influences attitudes toward health care technology [55,56]. Thus, 3 hypotheses about PEOU, PU, and attitude were formed:

- Hypothesis 1: PEOU has a positive impact on the PU of video consultations for type 2 diabetes care.
- Hypothesis 2: PEOU has a positive impact on attitudes toward video consultations for type 2 diabetes care.
- Hypothesis 3: PU has a positive impact on attitude toward video consultations for type 2 diabetes care.

The BI to use represents an individual's intention to use a new technology [41]. BI to use is an important component as it is a proxy capable of predicting subsequent actual user behavior in health care and beyond [33,41,42]. According to TAM, the extent to which users perceive a technology to be useful is directly influenced by their ratings of BI to use [38]. In the context of general practice, research has found a positive relationship between PU and BI to use [35,57-59]. Similarly, TAM suggests that the attitude of a user manifests itself as a positive or negative view of the BI to use technology. Research in the domain of primary health care finds that attitude influences the BI to use health care technology [23,60,61]. Thus, 2 hypotheses about PU, attitude, and BI to use were formulated:

- Hypothesis 4: PU has a positive impact on the BI to use video consultations for type 2 diabetes care.
- Hypothesis 5: Attitude toward video consultations for type 2 diabetes care has a positive impact on the BI to use the technology.

Figure 1. Research model based on the technology acceptance model.



Methods

Research Design and Setting

Data were collected through a cross-sectional web-based survey distributed to all general practitioners in Denmark (n=3326). The Danish health care system is mostly tax financed, and citizens can receive care from general practice free of per service charge. Danish general practitioners are self-employed but work on contracts for the public funder. Most general practitioners work in partnership practices, and their income is generated as a combination of fee for service and capitation [62]. The incentive for Danish general practitioners to use video consultations increased during the COVID-19 pandemic because of an agreement between the General Practitioners' Organization (negotiating on behalf of Danish general practitioners) and the Danish Regions (responsible for procuring health services), which agreed on a fee for service to general practitioners to provide video consultations to patients.

Survey Measures

The main measures (13 items) central to our hypotheses originated from TAM [39] and health care studies [55] to ensure the validity of the measures. The measures were adapted to the specific context of general practice and video consultations, translated into Danish, and repeatedly examined to ensure consistency. PU, attitude, and BI to use were measured using 3 items each, and PEOU was measured using 4 items (Textbox 1). An item each in the attitude and BI to use dimensions was negatively worded to reduce the risk of agreement bias [55]. All items were measured on 5-point Likert scales, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). For PEOU, the items were worded according to the user status of the respondent (user vs nonuser of video consultations) to make the formulation relevant to the respondent. Respondents were able to skip questions or choose *do not know* (the latter being treated as missing data in subsequent analyses). Demographic measures (12 items) such as age and sex were collected to analyze the representativeness of the study sample in comparison with the total population of general practitioners. Before distribution and to test face validity, the survey was evaluated and revised according to inputs from 5 general practitioners working in each of the 5 Danish Regions.

Textbox 1. Items used in the research model.

- Perceived usefulness (PU)
  - PU1: can *improve* my treatment
  - PU2: can make my treatment more *effective*
  - PU3: can make my treatment *easier*
- Perceived ease of use (PEOU; worded differently for nonusers of video consultations as illustrated in brackets)
  - PEOU1: *learning* to use was (would be) easy
  - PEOU2: *(would be) easy to get* software to do what I need
  - PEOU3: (would be) easy to *master*
  - PEOU4: (would be) easy to *use*
- Attitude (ATT)
  - ATT1: using is a *good idea*
  - ATT2: using is *unpleasant*
  - ATT3: using is *beneficial*
- Behavioral intention (BI)
  - BI1: intend to use as *often as possible*
  - BI2: even when possible, *do not intend* to use
  - BI3: would *use to the extent possible*

Recruitment and Data Collection

The survey was administered using SurveyXact (Rambøll Management) [63]. To identify general practices, a list of all 1718 general practices in Denmark was obtained from MedCom (a provider of Danish public health care systems) [64] in January 2021. Of these 1718 practices, 44 (2.56%) general practices were excluded as they were managed by parties outside the target group of our study (eg, by Danish Regions). In total, 1674 general practices, representing 3326 general practitioners, were available for distribution [65].

The survey was distributed to general practices as an electronic letter on January 7, 2021, via the Danish public electronic mailbox system (e-Boks Business) using publicly available data from MedCom. The letter contained information about the study and a survey link. Participants were informed about data protection measures, anonymity of participation, and the option to be paid—DKK 276.72 (US $44) based on a General Practitioners'' Organization tariff—for the 20 minutes it maximally takes to complete the survey. The letter was addressed to the clinic, and all trained general practitioners were encouraged to participate. Unfortunately, it was not possible to contact each general practitioner directly as this information was not publicly available. The survey link was open and only available in a letter to ensure anonymity and availability for all general practitioners in a clinic. Data entry for payments was conducted in a separate survey to preserve anonymity. Two reminders were sent on January 21, 2021, and February 2, 2021. The data collection ended on February 7, 2021.

The Committee of Multipractice Studies in General Practice (journal number 25-2020) evaluated the study and recommended that general practitioners participate in the survey. This study was reported to the Danish Data Protection Agency (journal number 1-16-02-343-20).

Ethics Approval

The Research Ethics Committees for Central Denmark Region (1-10-72-181-20) concluded that the study could be conducted without approval from the committee as "According to the Consolidation Act on Research Ethics Review of Health Research Projects, Consolidation Act number 1083 of 15 September 2017, section 14(2) notification of questionnaire surveys or medical database research projects to the research ethics committee system is only required if the project involves human biological material."

Data Analysis

Data were analyzed using Stata (version 17.0; StataCorp) [66]. To compare sample demographics with the population of general practitioners, we analyzed the latter using registry data made available by the Danish Health Data Authority [67]. The measures used in TAM were analyzed for normality distribution, internal consistency, convergent validity, and discriminant validity. Normality was examined by calculating skewness, kurtosis, and the Mardia multivariate kurtosis test. Internal consistency was assessed using Cronbach α with an acceptable threshold of .70 [68]. Confirmatory factor analysis was performed to determine model validity. Factor loadings of ≥0.7 were deemed acceptable [69]. Subsequently, we explored the research model using structured equation modeling [70], which is standard in the data analysis of TAM [37]. We used quasi-maximum likelihood as the estimator, with Satorra-Bentler adjustments because of our findings of nonnormality for some of the measures [71]. *P*<.05 was set as the threshold for statistical significance.

We report the unstandardized and standardized path coefficients from structured equation modeling. The unstandardized path coefficients reflect the expected (linear) change in the dependent variable with each unit change in the independent variable, given the other variables in the model. The standardized path coefficients express relationships in the same unit; that is, SDs. The interpretation is that when an independent variable (eg, PU) changes by 1 SD, then the dependent variable (eg, BI to use) changes by an SD as well. By placing all coefficients in the same unit, the SDs for different variables measured in different metrics become interpretationally equivalent.

Results

Demographic Characteristics

A total of 457 general practitioners answered the survey, from which 32 (7%) incomplete responses were excluded, resulting in 425 (93%) respondents. The sample represented 12.78% (425/3326) of all Danish general practitioners. The sample represented 18.82% (315/1674) of Danish general practices. Compared with the population of general practitioners, Pearson chi-square tests showed that the individual characteristics of the study sample (ie, sex and age groups) were representative of the population not participating (Table 1). The sample differed with regard to general practice characteristics (ie, clinic and municipality type) as general practitioners from more partnership practices participated than from solo practices, and a larger share of general practitioners working in practices in the capital area participated. The incomplete responses had similar demographics to the complete responses, with most (23/32, 72%) dropping out during or directly after the demographic items.

Table 1. Overview of respondents in sample and comparison with the remaining population.

Characteristics[a]	Survey sample (n=425), n (%)	Population not in the sample (n=2901), n (%)	Pearson chi-square (*df*)
Sex (female)[b]	226 (53.1)	1659 (57.1)	0.2 (1)
Age group (years)[b]			0.8 (6)
30-39	26 (6.3)	205 (7.1)	
40-44	75 (18.1)	577 (20)	
45-49	100 (24.2)	614 (21.2)	
50-54	59 (14.3)	416 (14.4)	
55-59	64 (15.5)	433 (15)	
60-64	57 (13.8)	387 (13.4)	
≥65	33 (8)	260 (9)	
Municipality type where general practitioners work[c,d]			0.0 (4)
Capital area	133 (31.3)	789 (25.5)	
Large city	63 (14.8)	392 (12.7)	
Province city	88 (20.7)	754 (24.4)	
Suburban	70 (16.5)	507 (16.4)	
County	71 (16.7)	654 (21.1)	
Clinic type[c]			<0.001 (2)
Solo clinic	105 (25.1)	447 (35.7)	
Cooperation clinic	52 (12.4)	145 (11.6)	
Partnership clinic	419 (98.5)	659 (52.7)	

[a] Missing data in the population not in the sample and in the survey sample means that sums do not add to the population of general practitioners (N=3326), general practices (N=1674), and study sample (N=425).

[b] Population data from General Practitioners' Organization [65].

[c] Population calculated from data by the Danish Health Data Authority [67].

[d] Municipality types based on the definition by Statistics Denmark [72].

Measurements Based on the TAM

Table 2 presents the mean values (SD) of the 4 dimensions and the items from TAM. On a 5-point Likert scale, the highest mean value was PEOU 3.76 (SD 0.86) and ATT 3.48 (SD 0.92), thus indicating that respondents were confident that they, for instance, can use video consultations to manage type 2 diabetes and that the technology was a good idea. The mean values for PU 2.99 (SD 0.96) and BI to use 3.06 (SD 1.04) were similar, and the answers averaged around neither agreeing nor disagreeing. Across the studied dimensions and items, the data variability around the mean of the study sample was approximately 1 point on a 5-point Likert scale.

Table 2. Means and internal consistency of items in the research model (N=425).

Item	Participants, n (%)	Values, mean (SD)	Cronbach α
PU[a]			
PU1: can *improve* my treatment	389 (91.5)	2.70 (0.97)	.86
PU2: can make my treatment more *effective*	397 (93.4)	3.01 (1.07)	.78
PU3: can make my treatment *easier*	396 (93.2)	3.24 (1.13)	.85
PU: all usability items	379 (89.2)	2.99 (0.96)	.88
PEOU[b]			
PEOU1: *learning* to use was (would be) easy	417 (98.1)	3.99 (0.95)	.85
PEOU2: (would be) *easy to get* software to do what I need	401 (94.4)	3.81 (0.98)	.84
PEOU3: (would be) easy to *master*	412 (96.9)	3.91 (0.91)	.83
PEOU4: (would be) easy to *use*	372 (87.5)	3.28 (1.1)	.92
PEOU: all ease of use items	359 (84.5)	3.76 (0.86)	.89
ATT[c]			
ATT1: using is a *good idea*	409 (96.2)	3.29 (1.15)	.63
ATT2: using is *unpleasant*	398 (93.6)	2.04 (0.96)	.92
ATT3: using is *beneficial*	397 (93.4)	3.13 (1.09)	.68
ATT: all attitude items[d]	380 (89.4)	3.48 (0.92)	.83
ATT1+3: ATT excluding ATT2	393 (92.5)	3.21 (1.08)	.92
BI[e] to use			
BI1: intend to use as *often as possible*	403 (94.8)	2.66 (1.12)	.82
BI2: even when possible, *do not intend* to use	404 (95.1)	2.61 (1.2)	.88
BI3: would *use to the extent possible*	402 (94.6)	3.12 (1.12)	.78
BI: all intention items[f]	383 (90.1)	3.06 (1.04)	.88

[a] PU: perceived usefulness.

[b] PEOU: perceived ease of use.

[c] ATT: attitude.

[d] The mean represents all ATT variables with ATT2 reversed because of its negative wording.

[e] BI: behavioral intention.

[f] The mean represents all BI variables with BI2 reversed because of its negative wording.

The internal consistency of the items that comprise the 4 dimensions in TAM had Cronbach α >.8 (Table 2). Cronbach α values of ≥.7 indicate acceptable internal consistency. Although the internal consistency of attitude was .83, this value should be interpreted with caution. The right-hand column of Table 2 shows the effect of removing 1 of the 3 items on Cronbach α; that is, for the attitude dimension, the Cronbach α drops to .63 and .68 when removing items 1 and 2 and increases to .92 when removing item 3. In addition to attributing this change in internal consistency to this analytical finding, free-text remarks by some respondents indicated that the negative wording of item 3 could be confusing and challenging to answer. On the basis of logical reasoning [73] and to reflect the attitude dimension more accurately, we excluded item 2 from the subsequent analysis.

To determine the correct structural equation modeling estimation method, we calculated the skewness and kurtosis of all the measures to examine normality. The results showed a mild degree of skewness (ranging from −0.971 to 0.232) with moderate kurtosis (ranging from 2.134 to 3.841). Normality was further evaluated using the Mardia multivariate kurtosis test, in which all dimensions failed except attitude, thereby indicating nonnormally distributed measures (PU 20.4, χ^2_1=90.9, *P*<.001; PEOU 43.3, χ^2_1=694.6, *P*<.001; attitude 8.22, χ^2_1=0.3, *P*=.57; BI 17.9, χ^2_1=26.0, *P*<.001). As nonnormality invalidates the assumption for the maximum likelihood method of structural equation modeling estimation, we used Satorra-Bentler adjustments to relax the assumption of normality. The measures in TAM were also assessed for convergent validity and discriminant validity (Table 3).

The measures were further validated using a confirmatory factor analysis that showed factor loadings >0.7, except for the item PEOU4—*easy to use* (0.63). PEOU4 was also an outlier in terms of missing data, with 12.7% (53/425) of missing responses, leading to the suspicion that the data were not missing at random. We excluded PEOU4 from the analysis and ran a new confirmatory factor analysis, which had factor loadings ranging from 0.77 to 0.92, thereby confirming that the latent variables of TAM were explained by the observed variables. Goodness-of-fit indices confirmed that the confirmatory factor analysis was a good fit for the data (χ^2_{38}=51.5, χ^2/df=1.4; *P*=.07; root mean squared error of approximation 0.033 [recommended value <0.05]; standardized root mean square residual 0.024 [recommended value <0.08]; comparative fit index 0.995 [recommended value >0.95]) [74]. The final research model included data from 76.9% (327/425) of respondents.

Table 3. Correlations between dimensions and items in the research model.

Item	PU[a]	PEOU[b]	ATT[c]	BI[d]
PU				
PU1	0.731	0.213	0.702	0.640
PU2	0.824	0.335	0.761	0.700
PU3	0.747	0.328	0.785	0.701
PEOU				
PEOU1	0.204	0.803	0.250	0.378
PEOU2	0.181	0.826	0.265	0.359
PEOU3	0.224	0.853	0.301	0.410
PEOU4	0.477	0.607	0.553	0.551
ATT				
ATT1	0.800	0.419	0.844	0.789
ATT3	0.801	0.369	0.844	0.765
BI				
BI1	0.703	0.454	0.754	0.813
BI2	0.613	0.441	0.668	0.711
BI3	0.709	0.426	0.750	0.773

[a] PU: perceived usefulness.

[b] PEOU: perceived ease of use.

[c] ATT: attitude.

[d] BI: behavioral intention.

Hypothesis Testing

We used structural equation modeling to analyze our hypotheses and the final research model. The goodness-of-fit indices model showed an acceptable fit (Table 4).

Analysis of the research model using unstandardized coefficients (Figure 2; Table 5) showed that the original paths of the model were significant (*P*<.005), except for the path from PU to BI to use (*P*=.84). PEOU had a positive influence on PU (β=.26, 95% CI 0.14-0.38) and attitude (β=.16, 95% CI 0.08-0.24). PU had a positive influence on attitude (β=1.22, 95% CI 1.09-1.36). The influence of attitude and PU on BI to use was also positive (β=.82, 95% CI 0.52-1.12; β=.04, −0.38 to 0.47); however, the latter was statistically insignificant. The calculated R^2 values (Figure 2) showed that 82% of the variance in BI to use was explained by attitude and PEOU, with attitude having the strongest influence. Standardized coefficients showed similar results (Figure 2; Table 6) and indicated that the strongest relationship existed between PU and attitude and between attitude and BI.

Table 4. Fit indices for structural equation modeling estimation.

Fit index	Structural equation modeling model with Satorra-Bentler	Recommended value [74,75]
Chi-square (*df*)	63.59 (39)	N/A[a]
Chi-square/*df*	1.63	<3.0
P value>chi-square (*df*)	0.008	>0.05
Root mean squared error of approximation	0.044	<0.05
Comparative fit index	0.991	>0.95
Tucker-Lewis index	0.987	>0.95
Standardized root mean square residual	0.036	<0.08

[a] N/A: not applicable (the literature on structural equation modeling does not recommend a value).

Figure 2. Results of structural equation modeling, unstandardized (and standardized) coefficients. *$P<.001$.

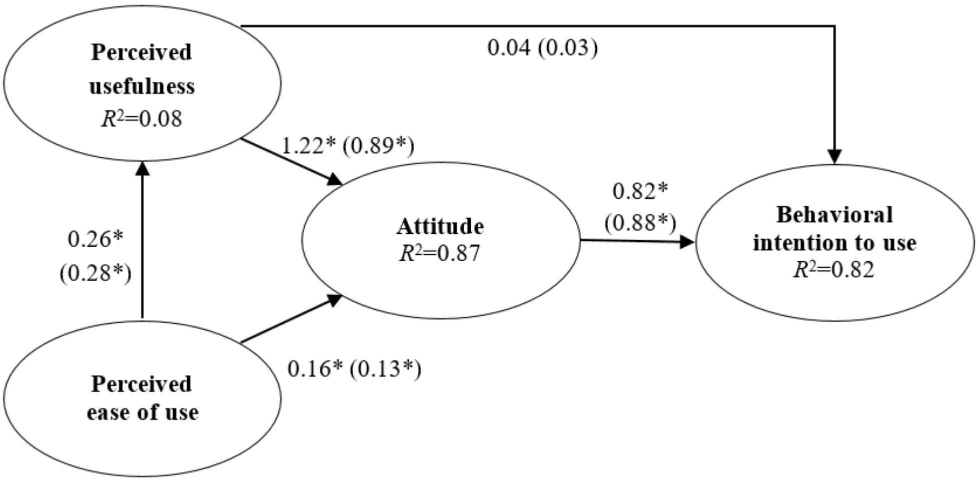


Table 5. Structural equation modeling estimation, unstandardized coefficients[a].

Path	β coefficient	*z* value	*P* value	95% CI
PEOU[b]→PU[c]	.26	4.26	<.001	0.14 to 0.38
PU→attitude	1.22	17.44	<.001	1.09 to 1.36
PEOU→attitude	.16	4.01	<.001	0.08 to 0.24
PU→BI[d]	.04	0.20	.84	−0.38 to 0.47
Attitude→BI	.82	5.35	<.001	0.52 to 1.12

[a] Satorra-Bentler adjusted; unstandardized coefficients.

[b] PEOU: perceived ease of use.

[c] PU: perceived usefulness.

[d] BI: behavioral intention.

Table 6. Structural equation modeling estimation, standardized coefficients[a].

Path	β coefficient	*z* value	*P* value	95% CI
PEOU[b]→PU[c]	.28	4.09	<.001	0.15 to 0.42
PU→attitude	.89	38.19	<.001	0.84 to 0.94
PEOU→attitude	.13	4.09	<.001	0.07 to 0.19
PU→BI[d]	.03	0.19	.85	−0.31 to 0.37
Attitude→BI	.88	5.54	<.001	0.57 to 1.19

[a] Satorra-Bentler adjusted; standardized coefficients.

[b] PEOU: perceived ease of use.

[c] PU: perceived usefulness.

[d] BI: behavioral intention.

Discussion

Principal Findings and Comparison With Prior Work

To explore the potential of using video consultations to provide type 2 diabetes care in general practice, we used insights from technology adoption [36-40] to systematically elicit the technology acceptance of general practitioners. From our survey of Danish general practitioners, we found support for 4 of the 5 research hypotheses (standardized and unstandardized path coefficients).

First, our findings suggest that PU is the primary driver of a positive attitude toward using video consultations to provide type 2 diabetes in general practice (hypothesis 3 accepted: unstandardized β=1.22, 95% CI 1.09-1.36). Similarly, earlier research in general practice found that this relationship appeared to be highly important [53,54]. The unstandardized path coefficient indicates that increasing the PU of the technology by 1 unit will increase the attitude by 1.22 units, given the other variables in the model. The standardized coefficient (β=.89, 95% 0.84-0.94) shows that a change of 1 SD in PU leads to an increase by 0.89 SDs in attitude. Second, attitude toward the technology is positively influenced by general practitioners' PEOU (hypothesis 2 accepted: unstandardized β=.16, 95% CI 0.08-0.24); however, the impact is lower than that for PU (β=1.22 vs β=.16). This finding mirrors previous studies that found that PU, not PEOU, is the primary driver of users' attitudes toward health care technology. A reason is that ease of use is not necessarily a sufficiently large benefit to offset the difficulties of integrating new technology into established work routines [76]. Another reason is that the importance of a technology that is easy to use tends to decrease with general technology use [38,55,56].

Third, our analysis confirmed the expectation that general practitioners' PU of video consultations would be positively influenced by their ratings of PEOU (hypothesis 1 accepted: unstandardized β=.26,95% CI 0.14-0.38). This mirrors findings from studies of other types of health care technology [50-52]. The relatively small impact of PEOU may be attributed to the high education level of Danish general practitioners who use IT technologies daily to deliver care, such as electronic patient records, and thus have a basic level of IT skills that could be speculated to give them confidence in learning new technologies. Fourth, the BI to use video consultations to provide type 2 diabetes was positively influenced by the attitude toward the technology (hypothesis 5 accepted: unstandardized β=.82, 95% CI 0.52-1.12). This particular relationship has also been found in other studies in the domain of primary health care [23,60,61]. Attitude is a central driver that corresponds to other influential theories of behavior change, such as the theory of planned behavior [77]. Fifth, our research model links PU to BI to use; however, the positive influence was statistically insignificant (hypothesis 4 rejected: unstandardized β=.04, −0.38 to 0.47). Compared with the impact of attitude, the influence of the PU of video consultations was also less influential (β=.82 vs β=.04). Studies from general practice generally report that PU has a positive influence on BI to use [35,57-59]. However, these studies do not include the attitude dimension from the original model [39] in their research models and, thus, do not address the relative importance. Our findings indicate that the BI to use video consultations for type 2 diabetes care is primarily the result of the positive impact PU has on attitude.

By studying chronic care in our context—type 2 diabetes—our research findings contribute to an emerging literature on video consultations in general practice that has hitherto mostly studied acute or subacute or out-of-hours care [17,25,31]. A major strength of the study is that the findings build on TAM, which is a robust model [36,37] with good predictive power for health technology use [38]. The findings are also supported by goodness-of-fit tests, showing that the research model has an acceptable fit for structural equation modeling estimation. A strength of our analysis is that it did not rely on the assumption that the measures were normally distributed as we used the Satorra-Bentler adjustments in the structural equation modeling.

Practical Implications

The potential of using video consultations in general practice to deliver chronic disease management is promising [1,10,11] and could fundamentally change how the primary care sector delivers care [12,19]. Type 2 diabetes is a chronic disease for which video consultations in general practice are particularly relevant [43-45] because, as a new care model, it can deliver high-quality, efficient care [46,47] at a time when the prevalence of diabetes is increasing [48]. Our findings (standardized and unstandardized path coefficients in the research model) indicate that the strongest positive relationships are between PU and attitude and between attitude and BI to use. This suggests that if a policy maker wants to increase general practitioners' use of video consultations to provide type 2 diabetes care, they must ensure that the technology is useful in general practice as it will have a positive influence on their attitude, which, in turn, will positively affect their intention to use the technology. Policy makers interested in scaling up video consultations could benefit from looking into the items of the dimensions that constitute the research model. For example, to improve PU, policy makers should find solutions to three questions: how can it be ensured that video consultations (1) improve treatment, (2) make treatment more effective, and (3) make treatment easier?

Relatedly, our findings provide suggestions for mitigating change hesitance, which remains a barrier to implementing video consultations in routine practice [21]. As research shows that working in larger practices—but not individual characteristics such as age or sex—increases the likelihood that a general practitioner uses video consultation [22,23], it appears relevant to explore the perceptions of small and large practices separately. Using the example of PU, small and large practices may differ in the ways in which video consultations can improve and make treatment easier. These insights are important as data from, for example, the Danish Health Authority show a decrease in the use of video consultations in general practice from 2020 to 2021 [78], which suggests that general practitioners use the technology but also that it is not yet a regular work routine in general practice. Moreover, continuous improvement of the technology and its use in practice is central as there is a risk that this new care model increases general practitioner workload, and there may be a need to allocate more resources to implement digital-first pathways [14]. To the latter end, research finds that training facilitates the implementation of video consultations in routine practice [20].

Limitations

Two modifications were made to the original TAM, underlining the final research model. First, an item (attitude item 2) was removed as it decreased the Cronbach α of the attitude dimension. Another item (PEOU4) was dropped because of the low factor loading from the confirmatory factor analysis. To assess the extent to which removing these items changed the findings, a structural equation modeling estimation, including these items, was performed, which showed path coefficients very similar to our final model, thereby supporting the accuracy of the final structural equation model. Second, structural equation modeling estimations were not performed with all respondents as those skipping questions were omitted. Running a structural equation modeling estimation that included respondents with missing answers resulted in similar path coefficients but had poorer goodness of fit. The final research model met the recommended indices of the goodness of fit but failed the chi-square test. Failing the chi-square test is a known issue with structural equation modeling, which, similar to our study, has a high number of respondents and survey answers that are not normally distributed [75]. The issue of nonnormality was addressed using Satorra-Bentler adjustments.

With the widespread research validation of TAM in combination with acceptable goodness-of-fit indices, the final research model is considered valid. However, as this study surveyed general practitioners from a tax-financed health care system, the findings may be most generalizable to countries with similar health care systems such as the English National Health System. Some authors also raise the concern that the original TAM and later extensions lack precision in health care because of their inability to consider the influence of external variables and barriers to technology acceptance [36] such as psychological ownership of IT [79] or social norms [55]. Nevertheless, for the purposes of this study, the research model was kept simple for 2 main reasons. First, findings from health care that extend TAM only result in a relatively modest increase in explanatory power [55]. Second, getting general practitioners to answer surveys is difficult [43], and including other variables to increase the precision a little would likely come at the expense of a lower response rate. More questions also increased the risk of respondent fatigue and missing answers.

The relatively low response rate of 12.8% of all 3326 Danish general practitioners increased the risk of selection bias. Nevertheless, it improved confidence in the findings that the individual characteristics of the sample of general practitioners were comparable with the population, and the share of respondents in the sample who used video consultations was similar to that of other sources [78]. This finding supports the generalizability of our results. The difficulty in getting Danish general practitioners to participate in survey research is an explanation as they operate as for-profit firms and are often on a tight schedule [62]. The survey was also distributed during the COVID-19 pandemic when other surveys of general practitioners had similar low response rates [22,43,80]. It could be speculated that general practitioners with the strongest positive or negative attitudes toward technology were more likely to participate. Univariate normality tests of the items in the attitude dimension, as mentioned previously, showed that the respondents' attitudes were relatively normally distributed and did not only represent the most negative or positive attitudes toward video consultations used for diabetes care.

The study design was cross-sectional and, thus, only capable of capturing the views of general practitioners at the time of data collection. Although the cross-sectional design is standard in most studies on TAM [37,38], longitudinal studies are generally recommended to assess changes over time to make study findings more robust. Collecting data on the variables in TAM from the same source (ie, general practitioners) makes common method bias [81] a potential risk in the study. However, common method bias is of modest importance here as the research model asks about the intention to use rather than actual use.

Conclusions

This study explored the potential of using video consultations to provide type 2 diabetes care in general practice by eliciting the technology acceptance of a representative survey sample of Danish general practitioners. On the basis of TAM, our study suggests 2 main drivers: PU positively affects attitude toward using video consultations for diabetes care, and attitude positively affects the BI to use the technology. For policy makers interested in scaling up general practitioners' use of video consultations to provide diabetes care, our findings indicate that they should emphasize how the technology can improve treatment and make it more effective and easier. To this end, policy makers may need to explore what these aspects of usefulness mean to general practitioners working in different organizational contexts.

Acknowledgments

The authors are grateful for the insights provided by the reference group of 5 general practitioners from 5 Danish regions: Rasmus Dahl-Larsen, Michel Kjeldsen, Gitte Krogh Madsen, Janus Nikolaj Laust Thomsen, and Mogens Vestergaard. The authors thank Henrik Støvring, Lasse Bjerg Hansen, Daniel Witte, and Kasper Norman for their methodological comments.

The authors would also like to thank the Quality and Training Committees of North Denmark Region, Central Denmark Region, Region of Southern Denmark, Region Zealand, the Capital Region, and the Lundbeck Foundation Scholarship for General Practice. Support was provided by the Steno Diabetes Center Aarhus, Aarhus University Hospital, which is partially funded by an unrestricted donation from the Novo Nordisk Foundation.

Authors' Contributions

TP and AS conceptualized and designed the study. TP and DCT collected the data, conducted the statistical analyses, and wrote the first draft of the manuscript. All the authors critically revised the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Litwin AS. Technological Change in Health Care Delivery: Its Drivers and Consequences for Work and Workers. UC Berkeley Labor Center. 2020 Jun. URL: https://laborcenter.berkeley.edu/wp-content/uploads/2020/07/Technological-Change-in-Health-Care-Delivery.pdf [accessed 2022-06-01]
2. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
3. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
4. An MH, You SC, Park RW, Lee S. Using an extended technology acceptance model to understand the factors influencing telehealth utilization after flattening the COVID-19 curve in South Korea: cross-sectional survey study. JMIR Med Inform 2021 Jan 08;9(1):e25435 [FREE Full text] [doi: 10.2196/25435] [Medline: 33395397]
5. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
6. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the nasss framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
7. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Deliv Res 2018 Jun;6(21):1-136. [doi: 10.3310/hsdr06210] [Medline: 29939519]
8. Tönnies J, Hartmann M, Wensing M, Szecsenyi J, Peters-Klimm F, Brinster R, et al. Mental health specialist video consultations versus treatment-as-usual for patients with depression or anxiety disorders in primary care: randomized controlled feasibility trial. JMIR Ment Health 2021 Mar 12;8(3):e22569 [FREE Full text] [doi: 10.2196/22569] [Medline: 33709931]
9. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021 Feb 25;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
10. Mold F, Cooke D, Ip A, Roy P, Denton S, Armes J. COVID-19 and beyond: virtual consultations in primary care-reflecting on the evidence base for implementation and ensuring reach: commentary article. BMJ Health Care Inform 2021 Jan;28(1):e100256 [FREE Full text] [doi: 10.1136/bmjhci-2020-100256] [Medline: 33436372]
11. Car J, Koh GC, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020 Oct 20;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
12. Salisbury C, Quigley A, Hex N, Aznar C. Private video consultation services and the future of primary care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]
13. Paré G, Raymond L, de Guinea AO, Poba-Nzaou P, Trudel MC, Marsan J, et al. Barriers to organizational adoption of EMR systems in family physician practices: a mixed-methods study in Canada. Int J Med Inform 2014 Aug;83(8):548-558. [doi: 10.1016/j.ijmedinf.2014.06.003] [Medline: 24969270]

14. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]
15. Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decis Support Syst 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
16. Rush KL, Howlett L, Munro A, Burton L. Videoconference compared to telephone in healthcare delivery: a systematic review. Int J Med Inform 2018 Oct;118:44-53. [doi: 10.1016/j.ijmedinf.2018.07.007] [Medline: 30153920]
17. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019 Apr 23;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
18. Randhawa RS, Chandan JS, Thomas T, Singh S. An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study. Prim Health Care Res Dev 2019 Jan;20:e5 [FREE Full text] [doi: 10.1017/S1463423618000361] [Medline: 29909798]
19. Wherton J, Shaw S, Papoutsi C, Seuren L, Greenhalgh T. Guidance on the introduction and use of video consultations during COVID-19: important lessons from qualitative research. BMJ Leader 2020 May 18;4(3):120-123. [doi: 10.1136/leader-2020-000262]
20. Hoffmann M, Wensing M, Peters-Klimm F, Szecsenyi J, Hartmann M, Friederich H, et al. Perspectives of psychotherapists and psychiatrists on mental health care integration within primary care via video consultations: qualitative preimplementation study. J Med Internet Res 2020 Jun 18;22(6):e17569 [FREE Full text] [doi: 10.2196/17569] [Medline: 32554369]
21. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
22. Scott A, Bai T, Zhang Y. Association between telehealth use and general practitioner characteristics during COVID-19: findings from a nationally representative survey of Australian doctors. BMJ Open 2021 Mar 24;11(3):e046857 [FREE Full text] [doi: 10.1136/bmjopen-2020-046857] [Medline: 33762248]
23. Jiwa M, Meng X. Video consultation use by Australian general practitioners: video vignette study. J Med Internet Res 2013 Jun 19;15(6):e117 [FREE Full text] [doi: 10.2196/jmir.2638] [Medline: 23782753]
24. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video consultations between patients and clinicians in diabetes, cancer, and heart failure services: linguistic ethnographic study of video-mediated interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]
25. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
26. Björndell C, Premberg Å. Physicians' experiences of video consultation with patients at a public virtual primary care clinic: a qualitative interview study. Scand J Prim Health Care 2021 Mar;39(1):67-76 [FREE Full text] [doi: 10.1080/02813432.2021.1882082] [Medline: 33650941]
27. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 May 1;4(1):bjgpopen20X101020 [FREE Full text] [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
28. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the acceptance of video consultation by patients in rural primary care: empirical comparison of preusers and actual users. JMIR Med Inform 2020 Oct 22;8(10):e20813 [FREE Full text] [doi: 10.2196/20813] [Medline: 32969339]
29. Bleyel C, Hoffmann M, Wensing M, Hartmann M, Friederich H, Haun MW. Patients' perspective on mental health specialist video consultations in primary care: qualitative preimplementation study of anticipated benefits and barriers. J Med Internet Res 2020 Apr 20;22(4):e17330 [FREE Full text] [doi: 10.2196/17330] [Medline: 32310139]
30. Chudner I, Drach-Zahavy A, Karkabi K. Choosing video instead of in-clinic consultations in primary care in Israel: discrete choice experiment among key stakeholders-patients, primary care physicians, and policy makers. Value Health 2019 Oct;22(10):1187-1196 [FREE Full text] [doi: 10.1016/j.jval.2019.05.001] [Medline: 31563262]
31. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
32. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
33. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]
34. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]

35. Saigí-Rubió F, Vidal-Alaball J, Torrent-Sellens J, Jiménez-Zarco A, López Segui F, Carrasco Hernandez M, et al. Determinants of Catalan public primary care professionals' intention to use digital clinical consultations (eConsulta) in the post-COVID-19 context: optical illusion or permanent transformation? J Med Internet Res 2021 May 31;23(6):e28944 [FREE Full text] [doi: 10.2196/28944] [Medline: 34097638]
36. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007 Dec;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
37. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
38. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
39. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
40. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
41. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
42. Sheppard BH, Hartwick J, Warshaw PR. The theory of reasoned action: a meta-analysis of past research with recommendations for modifications and future research. J Consum Res 1988 Dec;15(3):325-343. [doi: 10.1086/209170]
43. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
44. Verhoeven F, van Gemert-Pijnen L, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
45. 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]
46. Murphy ME, Byrne M, Galvin R, Boland F, Fahey T, Smith SM. Improving risk factor management for patients with poorly controlled type 2 diabetes: a systematic review of healthcare interventions in primary care and community settings. BMJ Open 2017 Aug 04;7(8):e015135 [FREE Full text] [doi: 10.1136/bmjopen-2016-015135] [Medline: 28780542]
47. Siegel KR, Ali MK, Zhou X, Ng BP, Jawanda S, Proia K, et al. Cost-effectiveness of interventions to manage diabetes: has the evidence changed since 2008? Diabetes Care 2020 Jul;43(7):1557-1592. [doi: 10.2337/dci20-0017] [Medline: 33534729]
48. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016 Apr 09;387(10027):1513-1530 [FREE Full text] [doi: 10.1016/S0140-6736(16)00618-8] [Medline: 27061677]
49. Larsson SC, Wallin A, Håkansson N, Stackelberg O, Bäck M, Wolk A. Type 1 and type 2 diabetes mellitus and incidence of seven cardiovascular diseases. Int J Cardiol 2018 Jul 01;262:66-70 [FREE Full text] [doi: 10.1016/j.ijcard.2018.03.099] [Medline: 29605469]
50. Venkatesh V, Zhang X, Sykes TA. “Doctors do too little technology”: a longitudinal field study of an electronic healthcare system implementation. Inf Syst Res 2011 Sep;22(3):523-546. [doi: 10.1287/isre.1110.0383]
51. Raymond L, Paré G, Ortiz de Guinea A, Poba-Nzaou P, Trudel MC, Marsan J, et al. Improving performance in medical practices through the extended use of electronic medical record systems: a survey of Canadian family physicians. BMC Med Inform Decis Mak 2015 Apr 14;15:27 [FREE Full text] [doi: 10.1186/s12911-015-0152-8] [Medline: 25888991]
52. Chaudhuri A, Saddikutti V, Prætorius T. iKure Techsoft: providing technology enabled affordable health care in rural India. Asian Case Res J 2018;22(02):385-411. [doi: 10.1142/s0218927518500165]
53. Krog MD, Nielsen MG, Le JV, Bro F, Christensen KS, Mygind A. Barriers and facilitators to using a Web-based tool for diagnosis and monitoring of patients with depression: a qualitative study among Danish general practitioners. BMC Health Serv Res 2018 Jun 27;18(1):503 [FREE Full text] [doi: 10.1186/s12913-018-3309-1] [Medline: 29945613]
54. Kivekäs E, Enlund H, Borycki E, Saranto K. General practitioners' attitudes towards electronic prescribing and the use of the national prescription centre. J Eval Clin Pract 2016 Oct;22(5):816-825. [doi: 10.1111/jep.12548] [Medline: 27121026]
55. Chau PY, Hu PJ. Investigating healthcare professionals’ decisions to accept telemedicine technology: an empirical test of competing theories. Inf Manag 2002 Jan;39(4):297-311. [doi: 10.1016/S0378-7206(01)00098-2]
56. Hu PJ, Chau PY, Sheng OR, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manag Inf Syst 1999;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
57. Asua J, Orruño E, Reviriego E, Gagnon MP. Healthcare professional acceptance of telemonitoring for chronic care patients in primary care. BMC Med Inform Decis Mak 2012 Nov 30;12:139 [FREE Full text] [doi: 10.1186/1472-6947-12-139] [Medline: 23194420]

58. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jan;32(1-2):29-36. [doi: 10.1017/S0266462316000015] [Medline: 27472158]
59. Heselmans A, Aertgeerts B, Donceel P, Geens S, Van de Velde S, Ramaekers D. Family physicians' perceptions and use of electronic clinical decision support during the first year of implementation. J Med Syst 2012 Dec;36(6):3677-3684. [doi: 10.1007/s10916-012-9841-3] [Medline: 22402980]
60. Pikkemaat M, Thulesius H, Milos Nymberg V. Swedish primary care physicians' intentions to use telemedicine: a survey using a new questionnaire - physician attitudes and intentions to use telemedicine (PAIT). Int J Gen Med 2021 Jul 15;14:3445-3455 [FREE Full text] [doi: 10.2147/IJGM.S319497] [Medline: 34295177]
61. Hung SY, Tsai JC, Chuang CC. Investigating primary health care nurses' intention to use information technology: an empirical study in Taiwan. Decis Support Syst 2014 Jan;57:331-342. [doi: 10.1016/j.dss.2013.09.016]
62. Pedersen KM, Andersen JS, Søndergaard J. General practice and primary health care in Denmark. J Am Board Fam Med 2012 Mar;25 Suppl 1:S34-S38 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110216] [Medline: 22403249]
63. SurveyXact. 2022. URL: https://www.surveyxact.com/ [accessed 2022-02-01]
64. Koder, tabeller, ydere. MedCom. 2021. URL: https://www.medcom.dk/opslag/koder-tabeller-ydere [accessed 2021-07-22]
65. Læge- og praksispopulationen, 1977-2020: Nøgletal fra medlemsregisteret. Praktiserende Lægers Organisation. 2020. URL: https://www.laeger.dk/sites/default/files/laege-_og_praksispopulationen_2020_004.pdf [accessed 2021-07-22]
66. Stata 17.0. StataCorp. 2022. URL: https://www.stata.com [accessed 2022-02-01]
67. Læger og tilmeldte patienter i almen praksis 2014-2020. Sundhedsdatastyrelsen. 2020. URL: https://sundhedsdatastyrelsen.dk/da/tal-og-analyser/analyser-og-rapporter/almen-praksis-og-kommuner/almen-praksis [accessed 2021-07-22]
68. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
69. Knekta E, Runyon C, Eddy S. One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. CBE Life Sci Educ 2019 Mar;18(1):rm1 [FREE Full text] [doi: 10.1187/cbe.18-04-0064] [Medline: 30821600]
70. Beran TN, Violato C. Structural equation modeling in medical research: a primer. BMC Res Notes 2010 Oct 22;3:267 [FREE Full text] [doi: 10.1186/1756-0500-3-267] [Medline: 20969789]
71. Tomarken AJ, Waller NG. Structural equation modeling: strengths, limitations, and misconceptions. Annu Rev Clin Psychol 2005;1:31-65. [doi: 10.1146/annurev.clinpsy.1.102803.144239] [Medline: 17716081]
72. Municipality Groups v1:2018. Statistics Denmark. 2018. URL: https://www.dst.dk/en/Statistik/dokumentation/nomenklaturer/kommunegrupper [accessed 2021-06-01]
73. Kopalle PK, Lehmann DR. Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. Organ Behav Hum Decis Process 1997 Jun;70(3):189-197. [doi: 10.1006/obhd.1997.2702]
74. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
75. Shi D, Lee T, Maydeu-Olivares A. Understanding the model size effect on SEM fit indices. Educ Psychol Meas 2019 Apr;79(2):310-334 [FREE Full text] [doi: 10.1177/0013164418783530] [Medline: 30911195]
76. Anderson JG. Clearing the way for physicians' use of clinical information systems. Commun ACM 1997 Aug;40(8):83-90. [doi: 10.1145/257874.257895]
77. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
78. COVID-19: Monitorering af aktivitet i sundhedsvæsenet: Beskrivelse af udviklingen i aktivitet i sundhedsvæsenet under COVID-19 epidemien - 10. rapport. Sundhedsstyrelsen. 2021 Jun. URL: https://www.sst.dk/-/media/Udgivelser/2020/Corona/Monitorering/10_-monitoreringsrapport.ashx [accessed 2021-06-04]
79. Barki H, Pare G, Sicotte C. Linking IT implementation and acceptance via the construct of psychological ownership of information technology. J Inf Technol 2008 Dec 01;23(4):269-280. [doi: 10.1057/jit.2008.12]
80. Saint-Lary O, Gautier S, Le Breton J, Gilberg S, Frappé P, Schuers M, et al. How GPs adapted their practices and organisations at the beginning of COVID-19 outbreak: a French national observational survey. BMJ Open 2020 Dec 02;10(12):e042119 [FREE Full text] [doi: 10.1136/bmjopen-2020-042119] [Medline: 33268433]
81. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]

Abbreviations

BI: behavioral intention
PEOU: perceived ease of use
PU: perceived usefulness
TAM: technology acceptance model

Edited by R Kukafka; submitted 03.03.22; peer-reviewed by M Pikkemaat, H Thulesius; comments to author 14.06.22; revised version received 07.07.22; accepted 18.07.22; published 30.08.22

Please cite as:
Torp DC, Sandbæk A, Prætorius T
The Technology Acceptance of Video Consultations for Type 2 Diabetes Care in General Practice: Cross-sectional Survey of Danish General Practitioners
J Med Internet Res 2022;24(8):e37223
URL: https://www.jmir.org/2022/8/e37223
doi: 10.2196/37223
PMID:

Original Paper

Health Information Seeking From an Intelligent Web-Based Symptom Checker: Cross-sectional Questionnaire Study

Kimberly Arellano Carmona[1], MPH; Deepti Chittamuru[1], PhD; Richard L Kravitz[2], MD, MSPH; Steven Ramondt[3,4], PhD; A Susana Ramírez[1], MPH, PhD

[1]School of Social Sciences, Humanities and Arts, University of California, Merced, CA, United States

[2]University of California, Davis, Sacramento, CA, United States

[3]Department of Donor Medicine Research, Sanquin Research, Amsterdam, Netherlands

[4]Department of Communication Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
A Susana Ramírez, MPH, PhD
School of Social Sciences, Humanities and Arts
University of California
5200 Lake Rd
Merced, CA, 95343
United States
Phone: 1 209 228 4400
Email: sramirez37@ucmerced.edu

Abstract

Background: The ever-growing amount of health information available on the web is increasing the demand for tools providing personalized and actionable health information. Such tools include symptom checkers that provide users with a potential diagnosis after responding to a set of probes about their symptoms. Although the potential for their utility is great, little is known about such tools' actual use and effects.

Objective: We aimed to understand who uses a web-based artificial intelligence–powered symptom checker and its purposes, how they evaluate the experience of the web-based interview and quality of the information, what they intend to do with the recommendation, and predictors of future use.

Methods: Cross-sectional survey of web-based health information seekers following the completion of a symptom checker visit (N=2437). Measures of comprehensibility, confidence, usefulness, health-related anxiety, empowerment, and intention to use in the future were assessed. ANOVAs and the Wilcoxon rank sum test examined mean outcome differences in racial, ethnic, and sex groups. The relationship between perceptions of the symptom checker and intention to follow recommended actions was assessed using multilevel logistic regression.

Results: Buoy users were well-educated (1384/1704, 81.22% college or higher), primarily White (1227/1693, 72.47%), and female (2069/2437, 84.89%). Most had insurance (1449/1630, 88.89%), a regular health care provider (1307/1709, 76.48%), and reported good health (1000/1703, 58.72%). Three types of symptoms—pain (855/2437, 35.08%), gynecological issues (293/2437, 12.02%), and masses or lumps (204/2437, 8.37%)—accounted for almost half (1352/2437, 55.48%) of site visits. Buoy's top three primary recommendations split across less-serious triage categories: primary care physician in 2 weeks (754/2141, 35.22%), self-treatment (452/2141, 21.11%), and primary care in 1 to 2 days (373/2141, 17.42%). Common diagnoses were musculoskeletal (303/2437, 12.43%), gynecological (304/2437, 12.47%) and skin conditions (297/2437, 12.19%), and infectious diseases (300/2437, 12.31%). Users generally reported high confidence in Buoy, found it useful and easy to understand, and said that Buoy made them feel less anxious and more empowered to seek medical help. Users for whom Buoy recommended "Waiting/Watching" or "Self-Treatment" had strongest intentions to comply, whereas those advised to seek primary care had weaker intentions. Compared with White users, Latino and Black users had significantly more confidence in Buoy (P<.05), and the former also found it significantly more useful (P<.05). Latino (odds ratio 1.96, 95% CI 1.22-3.25) and Black (odds ratio 2.37, 95% CI 1.57-3.66) users also had stronger intentions to discuss recommendations with a provider than White users.

Conclusions: Results demonstrate the potential utility of a web-based health information tool to empower people to seek care and reduce health-related anxiety. However, despite encouraging results suggesting the tool may fulfill unmet health information needs among women and Black and Latino adults, analyses of the user base illustrate persistent second-level digital divide effects.

(J Med Internet Res 2022;24(8):e36322) doi: 10.2196/36322

KEYWORDS

health information seeking; health information; information seeking; information seeker; information behavior; artificial intelligence; medical information system; digital divide; information inequality; digital epidemiology; symptom checker; digital health; eHealth; online health information; user demographic; health information resource; health information tool; digital health assistant

Introduction

Background

The ever-growing amount of health information available on the web is increasing the demand for tools that provide personalized and actionable health information. In addition, patients avidly seek information to inform their own health care decisions, either directly or by verifying information discussed during professional consultations. The broad scope of web-based health information includes generic information obtained through web-based searches and decision aids and tools that deliver personalized advice based on information specific to users. Such tools include symptom checkers that provide users with a potential diagnosis after responding to a set of probes about their symptoms.

Web-based symptom checkers are becoming increasingly popular, and the emergence of the COVID-19 pandemic has increased interest in these tools [1]. However, only a few studies have examined how and why they are used [2-4]. The limited research on symptom checkers has found generally positive effects of their use; technologically sophisticated web-based triage systems may help reduce unnecessary visits to emergency rooms and overuse of antibiotics [4], make health care accessible in low-resource settings [5], and increase patient engagement [6]. However, although the potential for their utility is great, more research is needed on the actual use and effects of such tools.

Some studies have raised concerns about the potential of web-based health information systems to spread disinformation and inaccurate diagnostic information [2,7,8]. For example, a study evaluating the diagnostic and triage accuracy of 23 web-based symptom checkers found that physicians performed better than the symptom checker algorithms [4]. However, physicians made incorrect diagnoses in 15% of the cases. Although research suggests that symptom checkers may be less effective than physicians in terms of diagnostic accuracy, it might be more critical that symptom checkers provide recommended actions (eg, whether symptoms warrant a trip to the hospital). Therefore, it is important to understand the impact of symptom checkers on how patients seek care and respond to health care advice.

A significant potential contribution of web-based symptom checkers as triage systems may be to reduce the negative effects of the current overwhelming health information environment, such as the health information overload experienced by web-based health information seekers and their struggle to discern reliable information from misinformation. A web-based medical information system that addresses the abovementioned problems can help people better understand the potential causes of the symptoms they are experiencing, empower them to seek the right kind of help, and potentially reduce anxiety caused by the symptoms they are experiencing.

Users must be able to trust and follow their recommendations for web-based symptom checkers to make meaningful contributions. If web-based symptom checkers are not trusted, they are less likely to be adopted by users, thereby limiting their potential [9]. Moreover, users may be unclear about the technology behind web-based symptom checkers. Research suggests that web-based symptom checkers’ artificial intelligence (AI) systems are neither transparent nor comprehensible to users, which may undermine trust in such tools [10]. Nevertheless, despite hesitancy and concerns regarding the accuracy, AI-powered symptom checkers have been perceived as useful for diagnosis by users [11].

A large body of research on information seeking grounded in the uses and gratification frameworks [12] has examined how people use different media to fulfill or gratify various needs. Research in this tradition has characterized health information–seeking behaviors by sources (ie, web-based vs offline seeking [13,14]) or objective (ie, seeking for themselves vs others [15-17]). Multiple studies have confirmed that active information seekers from nonclinical sources, including the internet, are more likely to be White, female, and have relatively high levels of education and income [18-22]. Racial differences in health information–seeking, as well as confidence in information and trust in various sources, have been well documented. There may be different levels of trust and use of sources by racial groups, which can lead to disparities if inaccurate sources are used [23].

Research based on self-reported media use has established that deliberate information seeking from media, including the internet, has been associated with better health outcomes [24], increased engagement in prevention behaviors [25], and more positive patient-clinician interactions [26-28] and has also assisted individuals in coping with uncertainty [20]. Web-based health information–seeking before presenting to an emergency physician also has the potential to improve patient-provider interaction without negatively affecting adherence to treatment [29].

Despite this extensive body of research on information seeking and the importance of the internet and other “new” media as sources of medical information, the quality of the evidence for the effects has been limited. Most previous studies examining information seeking from nonclinical sources, including nearly all internet-based health information–seeking studies, are limited by their reliance on self-reports of individuals’ information-seeking behaviors and behavioral or psychosocial outcomes. Furthermore, most studies rely on generalized, non–time-bound health information–seeking behaviors (ie, “Have you ever looked for information about [a topic] from [a source]”), or ask about information seeking within a specific timeframe, but do not examine the content of the information retrieved or the recommendation provided. Thus, the next frontier in this line of research is one that links objective measures of information seeking—both sources and content—with clinical and psychosocial outcomes to understand how people use the information they seek and find from nontraditional sources.

Objectives

This study aimed to address the methodological limitations of prior information-seeking research and examine who seeks information from an intelligent web-based symptom checker and for what purpose, how users experience the tool, what they intend to do with the information, and predictors of intentions to follow the tool recommendations. The following research questions (RQs) guided this study:

1. RQ1: Who uses a web-based symptom checker?
2. RQ2: What drives users to use a web-based symptom checker?
3. RQ3: What were the web-based symptom checker’s recommendations?
4. RQ4: How do users perceive the web-based symptom checker?
5. RQ5: What is the relationship between perceptions of a web-based symptom checker and intention to follow recommended actions?

Methods

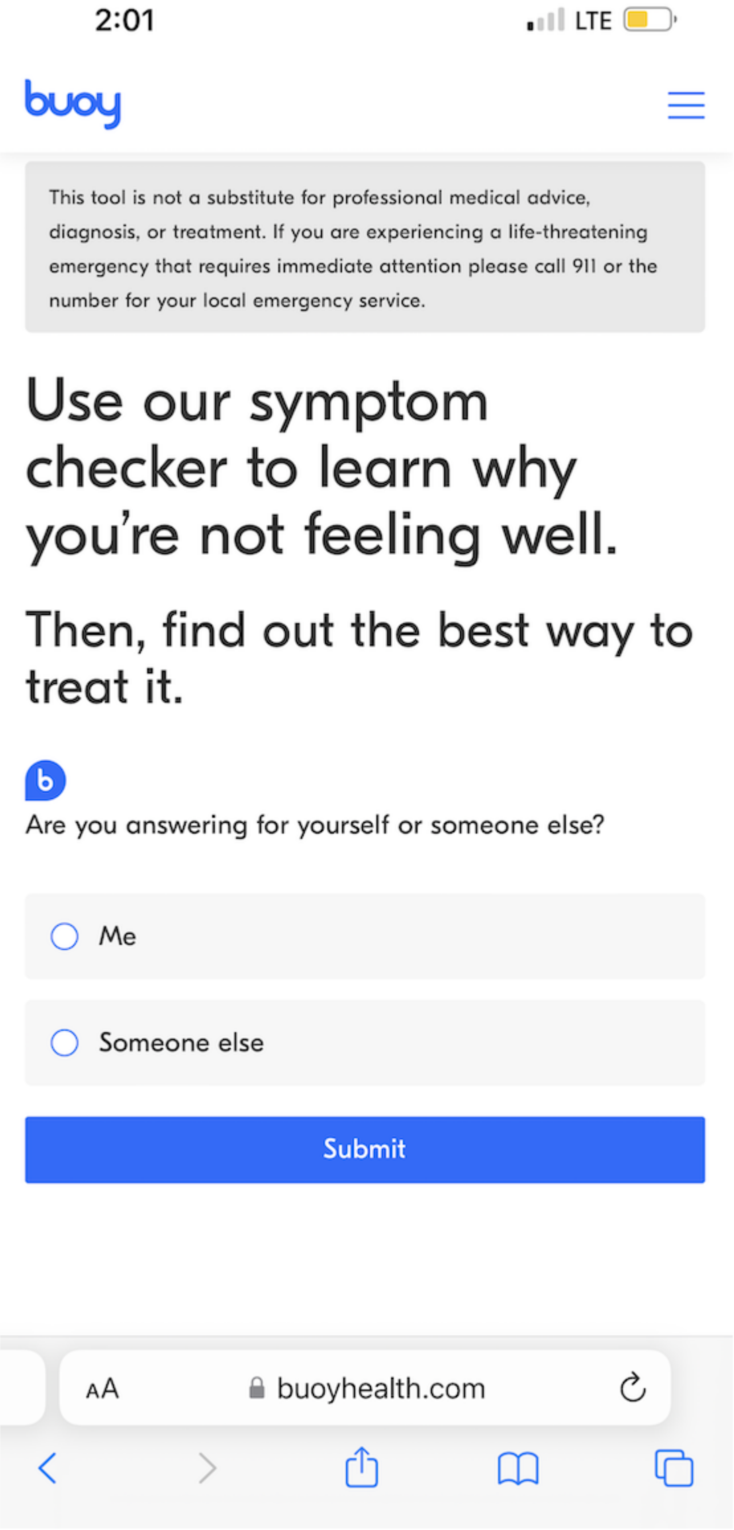
We conducted a cross-sectional survey of web-based health information seekers immediately following the completion of a visit to a web-based intelligent symptom checker, Buoy Health (Buoy Health, Inc [30]; N=2437).

Buoy Health: an AI-Powered Web-Based Symptom Checker

This cross-sectional study used data from patient encounters using Buoy Health, an AI-powered web-based symptom checker, between January 14, 2019, and February 28, 2019. Founded in 2014 by a team of physicians and researchers, the tool is based on conversational medical interviewing, mirroring a conversation with a provider. At the time of writing, Buoy’s symptom checker remains accessible for free on the web or through an app to any internet-connected person. The AI-powered tool uses a progressive series of health questions communicated via a chatbot to assess user symptoms (Figure 1). Buoy’s triage or diagnostic system by design offers health information customized for the user.

Buoy’s proprietary algorithm sources data from >18,000 clinical research studies [31]. Users receive 3 possible diagnoses and recommendations for appropriate levels of care (Figure 2). According to Buoy, the tool’s diagnostic accuracy is 90% [32]. Thus, tools such as Buoy—and other intelligent symptom checkers—have the potential to cut through the clutter of too much and contradictory information to provide personalized, science-based recommendations. A study examining how patients’ use of Buoy affected their plans for seeking care found that Buoy decreased uncertainty among users [33]. Buoy also lowered the level of urgency in patients associated with their condition. This study suggests tools such as Buoy are associated with users’ intended behavior when seeking care based on triage questions. Accordingly, our study adds to the growing literature that seeks to understand how patients use tools such as Buoy together with their providers to manage their health.

Figure 1. Screenshot of the patient-facing, artificial intelligence–assisted Buoy Health symptom checker.



This tool is not a substitute for professional medical advice, diagnosis, or treatment. If you are experiencing a life-threatening emergency that requires immediate attention please call 911 or the number for your local emergency service.

Use our symptom checker to learn why you’re not feeling well.

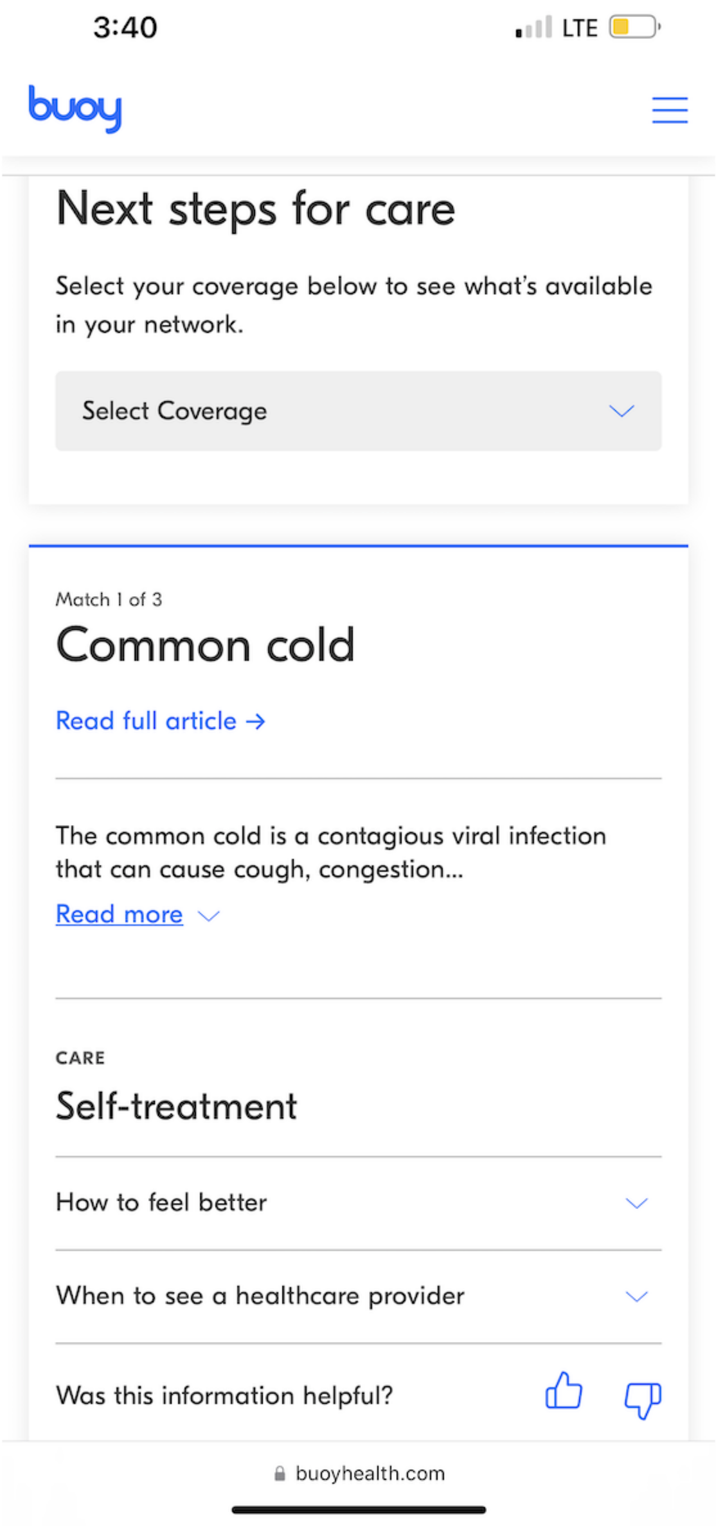
Then, find out the best way to treat it.

Are you answering for yourself or someone else?

○ Me

○ Someone else

Figure 2. Screenshot of a Buoy Health symptom checker recommendation.



Next steps for care

Select your coverage below to see what's available in your network.

Match 1 of 3

Common cold

The common cold is a contagious viral infection that can cause cough, congestion...

CARE

Self-treatment

Was this information helpful?

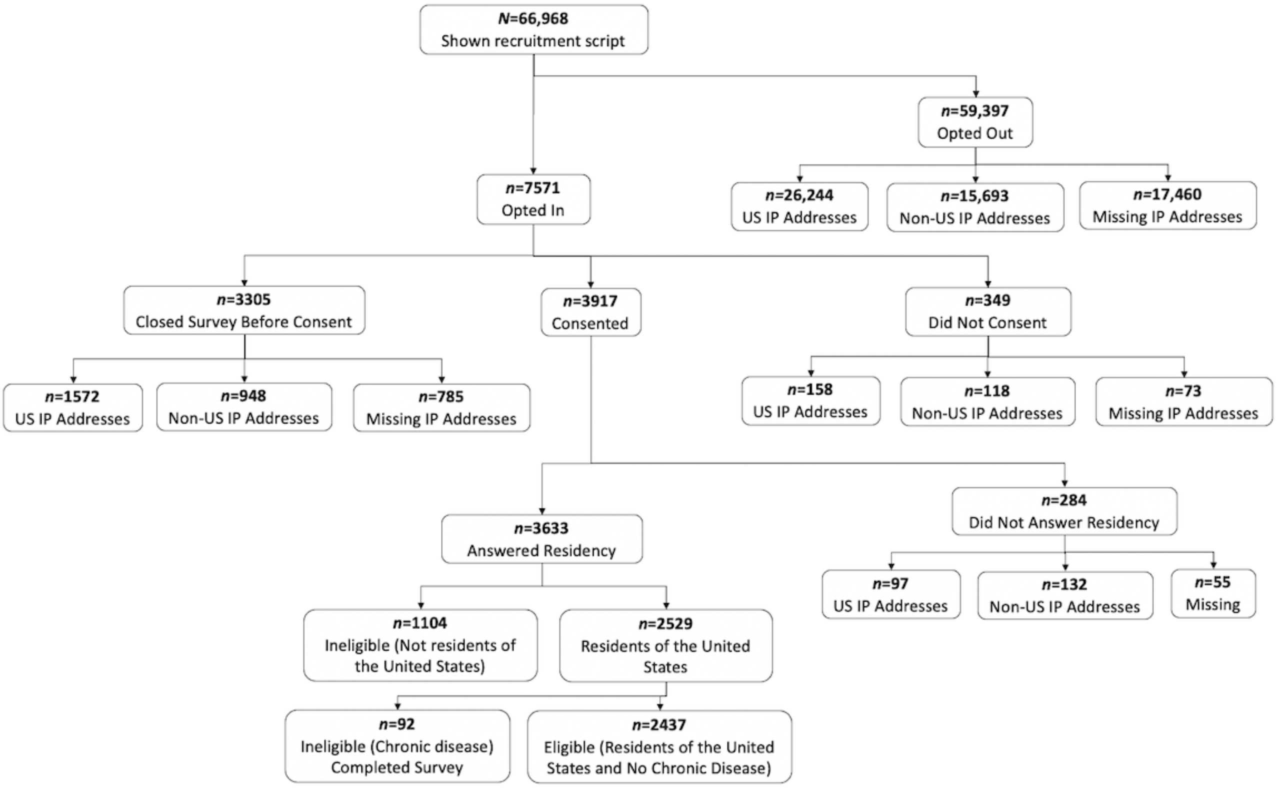
Sampling and Procedure

A recruitment script was shown to Buoy users, assumed to be web-based health information seekers, who met the inclusion criteria via a pop-up window immediately following their Buoy session. Inclusion criteria included completion of the Buoy interview to the recommendation stage in ≤10 minutes, being aged ≥18 years, and residency in the United States (although not necessarily physically in the United States at the time of seeking).

In addition, potential participants must have completed the Buoy interview for themselves; that is, they were seeking information about their symptoms. As 95% of Buoy users complete the diagnostic interview within 10 minutes, users who took longer were not representative of the typical user and thus were not invited to participate in our study to avoid other unanticipated ways in which they might differ from the typical user. Using similar logic, we excluded people who had a pre-existing serious or chronic condition [34] as they may not be representative of the typical Buoy user either. It is expected that their health information–seeking habits and use patterns of Buoy would be different from all other Buoy users. Finally, for ethical reasons, we excluded users who Buoy advised to seek immediate medical care from eligibility, including immediate medical care via 911 or in the emergency department. Figure 3 shows attrition at each stage.

Participants received a US $5 electronic gift card in appreciation of their time following completion of the survey, which had a mean time to completion of 8.61 (SD 6.78) minutes. The gift cards were delivered to an email address that was also used for follow-up. Participants were informed that they would receive another incentive (US $10) following the completion of a second survey. A follow-up assessment was administered 2 weeks after the initial survey to those who chose to provide an email address; however, this study reports only the baseline data.

Figure 3. Flow diagram showing attrition of participants.



Constructs and Measures

Overview

This study followed the tradition of uses and gratifications research [12]. We sought to understand who uses Buoy, perceptions of the user experience, and what they intend to do with the information they obtain. The survey was guided by the integrative model of behavior change [35]. The key constructs and measures are described in the following sections, and the complete survey instrument is available upon request from the corresponding author.

Reasons for Using Buoy

Patients could select ≥1 of 5 reasons for using Buoy; the list was based on the review of internet use for appraisal of symptoms of physical illness by Mueller et al [3]. Options included not being confident that the health care provider provided the correct diagnosis, symptoms not serious, sensitive or embarrassing symptoms, new symptoms, and persistent symptoms. An open-ended response was also provided, and the results were interpreted by 2 coders to map to original close-ended or new codes (access, anxiety, curiosity, and triage).

Trust in Health Information Sources

Trust in a variety of health information sources was assessed using a single Likert item, with responses ranging from “Not at all” (1) to “A lot” (4), adapted from the National Cancer Institute’s Health Information National Trends Survey [36]. The question stem was “In general, how much would you trust information about health or medical topics from each of the following?” The list of sources was randomized: physician/family or friends/newspapers or magazines/radio/internet news/television/government health agencies/social media (such as Facebook or Twitter)/Google/blogs/Buoy.

Prior research has demonstrated that the usability of the health information website affects trust in and credibility of the health information found on the site [37]. Thus, we assessed confidence, comprehensibility, perceived utility, and the emotional effects of using Buoy.

Confidence in Smart Symptom Checkers

A total of 2 items with 5-point response values from "Not at all confident" to "Very Confident" were adapted from Sivakumar et al [38] and combined as a scale where higher values represented greater confidence.

Comprehensibility of Smart Symptom Checkers

We assessed the extent to which the language on the website was easy to understand and the website was understandable and easy to read using 3 response values of 5 points (strongly disagree to strongly agree). Items were based on a scale by van Weert et al [39], with higher values representing greater comprehensibility of Buoy.

Perceived Utility of Smart Symptom Checkers

A total of 3 items with 5-point response values (strongly disagree to strongly agree) and combined as a scale by Davis [40] were used to assess the extent to which the website made the diagnosis of symptoms quicker and easier and the website's overall usefulness.

Emotional Effects of Using Smart Symptom Checkers

The emotional effect was informed by White and Horvitz [41] and was measured using a scale of 2 items of 5 points (strongly disagree to strongly agree). The items assessed anxiousness about a perceived medical condition and the extent to which the website encouraged help seeking. Higher values represented more positive emotional effects of using the website.

Recommended Action

At the conclusion of the Buoy diagnostic interview, users were given at least one and up to 3 of 8 possible recommendations for the next steps (triage level) that correspond to their cluster of symptoms and potential diagnosis: (1) wait and watch, (2) self-treatment, (3) phone call or in-person visit in the next 3 days, (4) primary care physician in 2 weeks, (5) primary care physician in 1 to 2 days, (6) in-person visit that day or as soon as possible, (7) hospital emergency room, and (8) emergency medical service. Buoy users who received the 2 most urgent recommended actions were not included in our study for ethical reasons. A complete description of the recommendations is available in Multimedia Appendix 1. Buoy provided the research team with the actual recommendations shown to all eligible users. In addition, we asked participants to indicate which of the 6 possible recommendations they had received from Buoy. We compared participants' self-reports with the Buoy-reported recommendations as a manipulation check. The comparison matched survey responses with at least one of the Buoy recommendations. Most self-reported recommendations matched at least one recommendation, as reported by Buoy (1595/2141, 74.49%).

Intention to Follow Recommended Action

The reasoned action approach informed this measure [35]. Intention was measured for all 6 included recommended actions and intentions to discuss Buoy's recommendation with a physician or other health professional. The response values ranged from 1 (strongly disagree) to 5 (strongly agree). Examples of the statement are as follows, "I intend to [follow Buoy's recommendation]" and "I intend to discuss the information I got from Buoy with my doctor or other health care professional." Higher values on the item represented a stronger intention to follow Buoy's recommendations or discuss the same with the physician. The recommended action was assessed as a binary variable. Users who scored 4 or 5 on intention (high) were classified as having medical intention, whereas those who scored 1, 2, or 3 were classified as having no medical intention. Intention to follow "Wait and Watch" and "Self-Treatment" were combined on a "No Medical Intention" scale. The intentions "Phone Call or In-Person Visit in the next 3 Days," "Primary Care Doctor in 1-2 days," and "In-Person Visit Today or ASAP" were combined in a "Medical Intention" scale.

Coding of Symptoms and Diagnoses

Users' self-reported symptoms resulting from the Buoy interview were coded into 13 categories using the Centers for Disease Control and Prevention National Ambulatory Medical Care Survey (NAMCS) coding protocol [42]. During the interview, the users were prompted to enter up to 5 presenting symptoms. We report only on the first as that was the primary issue driving the use of the web-based symptom checker. Using an iterative coding process, we generated a set of unique symptoms (N=2040) and unique diagnoses (N=938) from all Buoy data sets.

A total of 2 coders independently coded the first symptom. Coder 1 was part of the codebook development process. Coder 2 was introduced into the study once the codebook was finalized. Disagreements were resolved by discussion. The second author resolved disagreements when consensus was not possible. Cohen κ was run to determine interrater reliability between the 2 coders' assignment of NAMCS codes for the 2040 unique symptoms; there was substantial agreement between the 2 coders (κ=0.73; [43]). We further categorized whether the first symptoms were serious and likely to require medical attention based on Shapiro et al [44] (chest pain that is heart related, bleeding, loss of consciousness, shortness of breath, and weight loss).

Users are provided with up to 3 possible diagnoses or display names at the completion of their interview, ranked and weighted according to Buoy's proprietary algorithm, along with recommendations for subsequent actions. Diagnoses were coded into 25 categories comprising major systems, disorders, and conditions, in line with the NAMCS. We report the first diagnosis display name as the algorithm had the highest confidence in it. In addition, the first diagnosis display name had no missing data.

Analytic Approach

For this descriptive analysis, we computed frequencies and percentages to summarize participant characteristics and experiences, overall and by sex and ethnicity where relevant, and to assess intentions to comply with Buoy recommendations. ANOVAs with Bonferroni correction examined the mean outcome differences between racial or ethnic groups on user experience and recommendations. Nonparametric tests in the form of the Wilcoxon rank sum test were performed to test the mean differences between sexes in user experience and recommendations. The relationship between perceptions of the symptom checker and intention to follow recommended actions was assessed using logistic regression. Logistic regression models examined the factors affecting confidence in recommendations and intention to follow these. Analyses were conducted using R (version 4.0.3; R Foundation for Statistical Computing).

Ethics Approval

The University of California, Merced Institutional Review Board approved this study (approval number: UCM2018-124).

Results

Users of a Web-Based Symptom Checker

Consistent with prior studies on health information seekers, Buoy users were well-educated (1384/1704, 81.22% some college or more), mostly White (1227/1693, 72.47%), and female (2069/2437, 84.89%). The mean age of the users was 39.4 (SD 14.7) years. Users were similar to other users of web-based symptom tools, and a prior study of web-based symptom checkers found that users were predominantly female with a mean age of 40 years [33]. Findings from other studies further indicate an age, sex, and socioeconomic divide among adults' web-based health information–seeking behaviors [45]. The sampled users were also relatively privileged in terms of health care access; most had insurance (1449/1630, 89%) and a regular health care provider (1307/1709, 76%). They were generally in good health; 59% (1000/1703) reported their health as good, very good, or excellent. Table 1 shows the additional demographic details.

Table 1. Sample characteristics and comparison with all users of an intelligent web-based symptom checker.

Characteristics	Analytic sample[a] (N=2437)	Eligible opt-outs (N=27,816)
Age (years)		
Values, mean (SD)	39.35 (14.43)	36.92 (14.13)
Values, range	18-87	18-89
Ethnicity (N=1693), n (%)		
White	1227 (72.47)	—[b]
Black or African American	189 (11.16)	—
Latino or Hispanic	139 (8.21)	—
Asian or Pacific Islander	86 (5.08)	—
Other	52 (3.07)	—
Education (N=1704), highest level completed, n (%)		
High school or less	320 (18.78)	—
Some college	689 (40.43)	—
College	695 (40.79)	—
Household income (US $; N=1654), n (%)		
<20,000	304 (18.38)	
20,000-34,999	226 (13.66)	—
35,000-49,999	232 (14.03)	—
50,000-74,999	316 (19.11)	—
75,000-99,999	237 (14.33)	—
≥100,000	339 (20.50)	—
General health status (self-reported; N=1703), n (%)		
Excellent	63 (3.70)	—
Very good	288 (16.91)	—
Good	649 (38.11)	—
Fair	532 (31.24)	—
Poor	171 (10.04)	—
Have regular health care provider (N=1709), n (%)	1307 (76.48)	—
Have insurance (N=1630), n (%)	1449 (88.90)	—

[a] The number of Buoy users in the analytic sample was 2437; during the period of the study, there were a total of 27,816 potentially eligible users (aged ≥18 years, US IP address, those seeking for themselves, and who completed the Buoy interview in <10 minutes) who opted not to participate.

[b] Data not available.

Drivers for Users to Use a Web-Based Symptom Checker

Users selected ≥1 of the 5 stated reasons for using Buoy, as well as open-ended responses, which were coded into 5 new categories. Over one-third (839/2437, 34.43%) of the users indicated persistent symptoms as a reason for using Buoy, followed by new symptoms (767/2437, 31.47%), symptoms not serious (545/2437, 22.36%), sensitive or embarrassing symptoms (269/2437, 11.04%), and not confident that health care provider provided correct diagnosis (220/2437, 9.03%). Less common reasons included new categories/codes: curiosity (66/2437, 2.71%), access (36/2437, 1.48%), anxiety (16/2437, 0.66%), triage (29/2437, 1.19%), and other (30/2437, 1.23%; data not shown).

Recommendations of the Web-Based Symptom Checker

We report the patterns in symptoms and diagnoses in 2 ways. First, we report the frequencies of symptoms and diagnoses organized by the NAMCS Biological Systems associated with them (Multimedia Appendix 2 [42,44]). Second, we report the top 10 symptoms and diagnosis categories, overall and by sex and race/ethnicity (Table 2).

Table 2. Top 10 symptoms and diagnoses (sorted into major categories), overall and by sex and ethnicity (N=2437).

Symptoms and diagnoses	Overall	Female (n=2069)	Male (n=368)	White (n=1227)	Latino (n=139)	Black (n=189)	Asian or Pacific Islander (n=86)
Primary symptom, proportion							
Musculoskeletal pain; headache; other pain	0.35	0.34	0.42	0.40	0.31	0.21	0.21
Axial musculoskeletal pain	0.05	0.05	0.06	0.06	0.02	0.02	0.01
Muscle pain	0.06	0.06	0.08	0.07	0.06	0.03	0.03
Joint pain	0.08	0.08	0.09	0.09	0.07	0.03	0.07
Headache	0.03	0.03	0.02	0.03	0.02	0.02	0
Chest pain	0.02	0.02	0.03	0.02	0.03	0.02	0.01
Other pain	0.11	0.10	0.13	0.11	0.10	0.08	0.08
Gynecological problems	0.12	0.14	0.00	0.08	0.16	0.22	0.26
All masses, lumps, and tumors	0.08	0.07	0.13	0.09	0.07	0.10	0.09
Edema	0.05	0.05	0.05	0.05	0.06	0.05	0.03
Skin issues	0.05	0.04	0.09	0.05	0.06	0.08	0.08
Gastrointestinal problems	0.05	0.05	0.04	0.06	0.10	0.03	0.06
Impaired sensation	0.04	0.03	0.06	0.04	0.01	0.04	0.03
Urinary tract problems	0.03	0.03	0.02	0.03	0.03	0.04	0.03
Acute upper respiratory tract symptoms	0.03	0.04	0.01	0.03	0.05	0.02	0.03
Other	0.18	0.19	0.17	0.19	0.15	0.22	0.16
Primary diagnosis, proportion							
Musculoskeletal conditions	0.12	0.12	0.17	0.14	0.08	0.07	0.07
Musculoskeletal injuries	0.05	0.05	0.05	0.06	0.04	0.04	0.02
Gynecological conditions	0.12	0.15	0	0.09	0.17	0.22	0.20
Skin problems	0.12	0.11	0.17	0.12	0.09	0.15	0.14
Infectious diseases	0.12	0.13	0.10	0.13	0.19	0.09	0.13
Digestive conditions	0.07	0.07	0.08	0.08	0.07	0.05	0.06
Neurological conditions	0.07	0.07	0.08	0.09	0.08	0.05	0.10
Cancer and benign growths	0.05	0.04	0.05	0.04	0	0.02	0.01
Urination problems	0.03	0.04	0.02	0.03	0.04	0.04	0.05
Endocrinal problems and conditions	0.03	0.03	0.02	0.03	0.01	0.04	0.03
Heart related issues	0.02	0.02	0.03	0.02	0.01	0.03	0.01
Other	0.17	0.18	0.22	0.18	0.24	0.19	0.17

Only 3 types of symptoms—pain (855/2437, 35.08%), gynecological issues (293/2437, 12.02%), and masses or lumps (204/2437, 8.37%)—accounted for almost half (1352/2437, 55.48%) of the site visits. The top 3 symptoms entered by men included pain (154/368, 41.8%), masses or lumps (49/368, 13.3%), and skin issues (33/368, 8.9%), whereas the top 3 symptoms in women included pain (701/2069, 33.88%), gynecological issues (293/2069, 14.16%), and masses or lumps (155/2069, 7.49%). Pain, gynecological issues, and masses or lumps were also reported as the top 3 symptoms for White, Black, and Asian or Pacific Islander users. The top 3 symptoms in Latino users were pain (43/139, 30.9%), gynecological issues (22/139, 15.8%), and gastrointestinal problems (14/139, 10.1%). In comparison, Native Americans, who represented <1% of users, only entered five symptoms: pain (4/13, 31%), gynecological issues (4/13, 31%), skin issues (1/13, 8%), gastrointestinal problems (1/13, 8%), and impaired sensation (1/13, 8%).

Among the entire sample, major diagnoses were musculoskeletal (303/2437, 12.43%), gynecological (304/2437, 12.47%) and skin conditions (297/2437, 12.19%), and infectious diseases (300/2437, 12.31%). Comparably, the top 3 diagnoses reported by Buoy for men included musculoskeletal conditions (63/368, 17.1%) and skin conditions (62/368, 16.8%) and infectious diseases (37/368, 10%). The top 3 diagnoses for women included gynecological conditions (304/2069, 14.69%), infectious diseases (263/2069, 12.7%), and musculoskeletal conditions

(240/2069, 11.59%). The diagnoses based on race or ethnicity followed a similar pattern. White users also reported musculoskeletal conditions (177/1227, 14.42%), infectious diseases (163/1227, 13.28%), and skin conditions (148/1227, 12.06%) as the top 3 diagnoses. Latino, Black, and Asian or Pacific Islander users reported gynecological conditions, skin conditions, and infectious diseases as the top 3 diagnoses. Gynecological conditions were reported as the top diagnosis category by Black (42/189, 22.2%) and Asian or Pacific Islander (17/86, 20%) users, whereas Latino users reported infectious diseases (26/139, 18.7%) as the top diagnosis category.

Buoy's primary recommendation was more evenly split across the less-serious triage categories. Users self-reported primary care physicians in 2 weeks (754/2141, 35.22%), self-treatment (452/2141, 21.11%), and primary care in 1 to 2 days (373/2141, 17.42%) as the top 3 recommendations provided by Buoy, followed by wait and watch (339/2141, 15.83%). Only 5.74% (123/2141) and 4.67% (100/2141) of users self-reported phone calls or in-person visits in the next 3 days and in-person visits that day or as soon as possible, respectively. The recommendations reported by Buoy closely matched primary care physicians in 2 weeks (924/2437, 37.91%), self-treatment (552/2437, 22.65%), and primary care in 1 to 2 days (456/2437, 18.71%). Most users (2098/2437, 86.09%) had 2 recommendations. Approximately 71.85% (1751/2437) had 3 recommendations, as reported by Buoy.

Users' Perceptions of the Web-Based Symptom Checking Experience

Users generally reported high levels of confidence in Buoy (mean 3.47, SD 0.97), found it useful (mean 4.18, SD 0.81) and easy to understand (mean 4.64, SD 0.53), and said that Buoy made them feel less anxious (mean 3.60, SD 1.05) and more empowered to seek medical help (mean 3.75, SD 0.96) Compared with White users, Latino and Black users had significantly more confidence in Buoy (*P*<.05), and the former also found it significantly more useful (*P*<.05; Table 3). Consistent with prior studies on trust in web-based health information sources [46-48], physicians were the most trusted source. However, Buoy was trusted more (mean 3.68, SD 0.61) than any other nonmedical source, including government agencies (mean 2.85, SD 0.95), family (mean 2.64, SD 0.76), and Google (mean 2.52, SD 0.79).

Table 3. Buoy user experience and recommendations (N=2437).

Item	Overall	Male (n=368)	Female (n=2069)	White (n=1227)	Latino (n=139)	Black (n=189)	Asian or Pacific Islander (n=86)
Comprehensibility of Buoy, mean (SD)	4.64 (0.53)	4.61 (0.49)	4.65 (0.53)	4.67 (0.50)	4.68 (0.55)	4.67 (0.53)	4.60 (0.45)
Buoy website was understandable, mean (SD)	4.60 (0.61)	4.57 (0.54)	4.60 (0.62)	4.63 (0.57)	4.63 (0.67)	4.60 (0.63)	4.57 (0.50)
Buoy website was easy to read, mean (SD)	4.66 (0.56)	4.62 (0.52)	4.67 (0.57)[a]	4.68 (0.54)	4.69 (0.59)	4.71 (0.55)	4.64 (0.48)
Language used on the Buoy website was easy to understand, mean (SD)	4.68 (0.55)	4.65 (0.51)	4.68 (0.55)	4.70 (0.51)	4.71 (0.58)	4.70 (0.54)	4.59 (0.49)
Confidence in Buoy, mean (SD)	3.47 (0.96)	3.39 (0.89)	3.49 (0.99)	3.44 (0.96)[b]	3.69 (0.92)[b]	3.63 (1.04)	3.48 (0.88)
Confidence in diagnoses, mean (SD)	3.34 (1.05)	3.27 (0.97)	3.36 (1.06)	3.29 (1.05)[b,c]	3.58 (0.99)[b]	3.53 (1.11)[c]	3.35 (0.96)
Confidence in the recommendation, mean (SD)	3.60 (1.02)	3.52 (0.95)	3.62 (1.03)	3.60 (1.01)	3.79 (0.98)	3.73 (1.09)	3.60 (0.91)
Perceived utility of Buoy, mean (SD)	4.18 (0.81)	4.14 (0.77)	4.19 (0.82)	4.16 (0.80)[d]	4.43 (0.73)[d]	4.25 (0.86)	4.20 (0.76)
Buoy enabled me to diagnose my symptoms more quickly, mean (SD)	4.15 (0.85)	4.11 (0.81)	4.16 (0.86)	4.12 (0.84)[e]	4.45 (0.75)[e,f]	4.20 (0.92)[f]	4.19 (0.80)
Using Buoy made the diagnosis of my symptoms easier, mean (SD)	4.16 (0.86)	4.12 (0.81)	4.16 (0.87)	4.13 (0.85)[b]	4.38 (0.79)[b]	4.23 (0.91)	4.14 (0.81)
Overall, I found Buoy useful to diagnose my symptoms, mean (SD)	4.23 (0.86)	4.19 (0.83)	4.24 (0.86)	4.22 (0.85)[b]	4.47 (0.75)[b]	4.31 (0.89)	4.27 (0.77)
Emotional consequences of using Buoy, mean (SD)	3.68 (0.90)	3.65 (0.79)	3.68 (0.91)	3.65 (0.88)	3.76 (1.02)	3.72 (1.00)	3.76 (0.66)
Less anxious, mean (SD)	3.60 (1.05)	3.56 (0.94)	3.61 (1.07)	3.58 (1.04)	3.70 (1.15)	3.59 (1.16)	3.67 (0.79)
Encouraged to seek help, mean (SD)	3.75 (0.96)	3.74 (0.88)	3.76 (0.98)	3.73 (0.95)	3.83 (1.11)	3.86 (1.05)	3.84 (0.76)

[a] Significant difference between sex (*P*<.05).

[b] Significant difference between White and Latino users (*P*<.05).

[c] Significant difference between White and Black users (*P*<.05).

[d] Significant difference between White and Latino users (*P*<.001).

[e] Significant difference between White and Latino users (*P*<.001).

[f] Significant difference between Latino and Black users (*P*<.05).

Relationship Between Perceptions of a Web-Based Symptom Checker and Intention to Follow Recommended Actions

Overall, most users reported intentions to follow Buoy's recommendations (1428/1886, 75.71%) and discuss Buoy's recommendations with a physician or health care professional (1198/1830, 65.44%; Table 4). Users reported the strongest intention to follow Buoy's wait and watch recommendation (mean 4.38, SD 0.90), followed by self-treatment (mean 4.33, SD 0.93), in-person visit that day or as soon as possible (mean 4.17, SD 1.01), phone call or in-person visit in the next 3 days (mean 4.05, SD 1.05), primary care physician in 2 weeks (mean 3.92, SD 1.19), and primary care physician in 1 to 2 days (mean 3.68, SD 1.26).

Intention to discuss Buoy's recommendations was positively associated with having a regular provider (odds ratio [OR] 1.37, 95% CI 1.04-1.82), and an income >US $50,000 was negatively associated (OR 0.75, 95% CI 0.57-0.98; OR 66, 95% CI 0.48-0.91; Table 5). Users aged between 35 and 44 years (OR 1.51, 95% CI 1.13-2.03) and 45 and 64 years (OR 1.57, 95% CI 1.18-2.10) had better intentions of discussing recommendations than younger users (aged 18-34 years). Compared with White users, Latino (OR 1.96, 95% CI 1.22-3.25) and Black (OR 2.37, 95% CI 1.57-3.66) users had stronger intentions to discuss recommendations with a provider, and Black users were twice as likely to intend to do so. Confidence in Buoy (OR 1.54, 95% CI 1.34-1.76), perceived utility (OR 1.32, 95% CI 1.10-1.58), and anxiety reduction because of using Buoy (OR 1.43, 95% CI 1.24-1.63) were associated with higher intention to discuss Buoy's recommendations.

Overall, users had strong intentions to follow Buoy's recommendations, and users who self-reported very good or excellent health had the strongest intention to wait or watch or self-treat (OR 1.92, 95% CI 1.04-3.65; Table 5). Those who reported Buoy as easy to read and understand were 2.2 times (95% CI 1.21-4.14) more likely to intend to wait or watch or self-treat than those who reported lower comprehensibility for Buoy. Users with health insurance (OR 2.21, 95% CI 1.36-3.57) and a regular provider (OR 1.59, 95% CI 1.11-2.28) had the strongest intentions to seek care. Confidence in Buoy (OR 1.87, 95% CI 1.56-2.25) and anxiety reduction because of Buoy (OR 1.54, 95% CI 1.29-1.83) were also associated with a higher intention to seek care.

Table 4. Intentions to follow and discuss Buoy recommendations (N=2437).

Item	Overall	Male (n=368)	Female (n=2069)	White (n=1227)	Latino (n=139)	Black (n=189)	Asian or Pacific Islander (n=86)
Intentions to follow Buoy's recommendations (n=1886), n (%)	1428 (75.71)	187 (9.91)	1241 (65.8)	908 (48.14)	116 (6.15)	149 (7.9)	62 (3.29)
Wait and watch (n=283), n (%)	249 (87.9)	24 (9.6)	225 (79.5)	146 (51.6)	23 (8.1)	29 (10.2)	22 (7.8)
Self-treatment (n=385), n (%)	339 (88.1)	50 (14.7)	289 (75.1)	226 (58.7)	34 (8.8)	32 (8.3)	8 (2.1)
Phone call or in-person visit in the next 3 days (n=107), n (%)	81 (75.7)	14 (13.1)	67 (62.6)	49 (45.8)	7 (6.5)	8 (7.5)	3 (2.8)
Primary care physician in 2 weeks (n=688), n (%)	487 (70.7)	60 (12.3)	427 (62.1)	317 (46.1)	35 (5.1)	48 (7.0)	17 (2.5)
Primary care physician in 1 to 2 days (n=336), n (%)	205 (61.0)	29 (14.1)	176 (52.4)	137 (40.8)	9 (2.7)	22 (6.5)	10 (3.0)
In-person visit that day or as soon as possible (n=87), n (%)	67 (77.0)	10 (11.5)	57 (65.5)	33 (37.9)	8 (9.2)	10 (11.5)	2 (2.3)
Intentions to discuss Buoy's recommendations (n=1830), n (%)	1198 (65.46)	156 (8.52)	1042 (56.94)	758 (41.42)	109 (5.96)	150 (8.19)	51 (2.79)

Table 5. Intentions to follow Buoy's recommendations.

Predictors	Discuss Buoy's recommendations		No medical intention		Medical intention	
	OR[a]	*P* value	OR	*P* value	OR	*P* value
Intercept	0.02 (0.01-0.06)	<.001[b]	0.04 (0.00-0.75)	.03[b]	0.02 (0.00-0.11)	<.001[b]
Age 35 to 44 years	1. 51 (1.13-2.03)	.006[b]	0.66 (0.32-1.40)	.28	1.09 (0.74-1.60)	.67
Age 45 to 64 years	1.57 (1.18-2.10)	.002[b]	0.57 (0.26-1.26)	.16	1.07 (0.74-1.55)	.70
Age ≥65 years	1.31 (0.79-2.21)	.30	0.97 (0.28-4.08)	.96	1.12 (0.58-2.27)	.74
Female	0.86 (0.62-1.20)	.39	0.79 (0.31-1.80)	.59	1.00 (0.65-1.54)	.99
Black	2.37 (1.57-3.66)	<.001[b]	0.62 (0.27-1.57)	.23	1.49 (0.89-2.54)	.14
Latino	1.96 (1.22-3.25)	.007[b]	1.56 (0.48-7.12)	.50	1.38 (0.74-2.68)	.33
Asian or Pacific Islander	1.04 (0.62-1.74)	.99	0.79 (0.24-3.23)	.72	0.82 (0.43-1.64)	.57
Other ethnicities	1.56 (0.80-3.18)	.20	0.67 (0.18-3.39)	.58	0.94 (0.41-2.28)	.89
Have insurance	0.79 (0.52-1.18)	.25	0.74 (0.24-2.01)	.57	2.21 (1.36-3.57)	.001[b]
Have regular provider	1.37 (1.04-1.82)	.03	0.51 (0.21-1.14)	.12	1.59 (1.11-2.28)	.01[b]
General health status: very good or excellent	1.09 (0.86-1.38)	.50	1.92 (1.04-3.65)	.04[b]	0.95 (0.70-1.29)	.73
Some college	0.95 (0.68-1.38)	.77	1.03 (0.41-2.46)	.95	0.89 (0.57-1.38)	.61
College degree	0.73 (0.52-1.04)	.08	0.54 (0.21-1.29)	.18	0.69 (0.43-1.08)	.11
US $50,000-99,999	0.75 (0.57-0.98)	.03[b]	1.55 (0.76-3.18)	.22	1.20 (0.85-1.70)	.31
≥US $100,000	0.66 (0.48-0.91)	.01[b]	1.74 (0.76-4.17)	.20	0.92 (0.61-1.38)	.68
Comprehensibility of Buoy	1.19 (0.93-1.53)	.17	2.24 (1.21-4.14)	.01[b]	0.90 (0.65-1.22)	.49
Confidence in Buoy	1.54 (1.34-1.76)	<.001[b]	2.23 (1.61-3.14)	<.001[b]	1.87 (1.56-2.25)	<.001[b]
Perceived utility of Buoy	1.32 (1.10-1.58)	.002[b]	1.02 (0.63-1.62)	.93	1.12 (0.90-1.39)	.32
Emotional consequences of using Buoy	1.43 (1.24-1.63)	<.001[b]	1.02 (0.66-1.53)	.93	1.54 (1.29-1.83)	<.001[b]

[a] OR: odds ratio.

[b] Significant association.

Discussion

Principal Findings

This study sought to understand who uses web-based AI-powered symptom checkers and for what purposes. The demographic profile of Buoy users was similar to that described in other studies of web-based health information seekers, suggesting that older, marginalized groups continue to be digitally excluded. Consistent with data on internet-based health-seeking behaviors more generally [49], most Buoy users were middle-aged (or younger), female, and highly educated. More research is needed to better understand older adults' web-based health information–seeking behaviors and support their medical and health decisions [50]. Although a scoping review of articles examining AI-driven symptom checkers from various perspectives found that those who do not have access to health care services are more likely to use symptom checkers [51], Buoy users overwhelmingly reported having health insurance. This finding does not negate the possibility that users were motivated by financial considerations, as most contemporary health plans require an out-of-pocket copayment. Nevertheless, this suggests that other considerations such as convenience were also salient.

Along these lines, prior research has identified an association between stigmatizing conditions and the use of symptom checkers [51]. In this study, gynecological problems were among the top 3 symptom groups. Furthermore, across presenting symptoms or diagnoses, approximately 11.04% (269/2437) of the respondents were "too embarrassed" to seek in-person care. Taken together, these findings suggest that symptom checkers might be particularly useful for users affected by conditions considered personal, embarrassing, stigmatizing, not warranting the physician's attention, or requiring potentially uncomfortable or psychologically stressful physical examinations (such as pelvic examinations).

In examining the reasons for using the tool, approximately one-third of the respondents had persistent symptoms that failed to resolve spontaneously, another one-third had new symptoms, and the rest either thought they did not need professional attention or (as mentioned previously) were too embarrassed to seek care. Thus, some patients used the symptom checker because they had significant health-related concerns; some because they lacked sufficient concern to warrant in-person care; and some because they had issues with perceived quality, cost, or convenience of available care or simply wanted a second opinion. Symptoms that persist longer than expected have been identified as strong drivers of health-related anxiety and, thus, health care use [52]. At the same time, valuing convenience and lack of trust in the health care system (factors that may be particularly prominent among young people and racially and ethnically minoritized groups, respectively) have been associated with a lower propensity to use formal health care services [53].

Regarding the user experience, users had high levels of confidence in Buoy and found it useful. Moreover, users trusted Buoy more than any other nonmedical source. Perceived confidence, utility, and trust were associated with a stronger intention to discuss Buoy's recommendations with a physician. This finding is in line with a study examining patient perspectives on the usefulness of a symptom checker [11]. Most Buoy users found the tool useful for diagnosis, and most reported that they would use it again. Although the experiences of users who discussed recommendations with their physicians varied, most felt that physicians were open to discussing the results of the tool. This is an important finding, as users may not follow recommendations to seek care if they believe that acting on the advice of a symptom checker will be questioned or even belittled by their physician, regardless of their confidence in the tool.

This study ultimately advances the understanding of web-based health information–seeking behaviors and outcomes by linking objective measures of information seeking from a web-based AI-powered system with clinical and psychosocial outcomes. The results demonstrate the potential utility of an artificially powered web-based health information tool to improve outcomes for users. Symptom checkers have been described as a means of addressing the lack of access to physicians and reducing unnecessary office visits [4].

There is a lack of research on whether the use of symptom checkers translates into medical care–seeking behaviors [4,33]. Future research should examine the effects of such tools on medical care seeking, specifically how users interpret recommendations, whether the recommendations are followed, and how user responses vary among sociodemographic groups. For example, one might surmise that individuals with limited access to care or with prior negative health care experiences might be more likely to attend to, appreciate, and follow such recommendations than their more privileged counterparts. Although symptom checkers may empower users to make more informed decisions, they might paradoxically worsen health disparities if their use were less accessible to some groups. Currently, web-based symptom checkers are mostly available for free. As web-based symptom checker companies establish partnerships with employers and health insurance companies to ensure profits, not all users may be equally ready or able to pay for symptom checking.

Limitations and Strengths

We partnered with the Buoy technical and medical staff to sample the users. Owing to our partnership approach, we were able to obtain the specific symptoms reported by the participants as the primary reason for using Buoy, as well as the possible diagnoses identified by Buoy and Buoy's triage recommendation. This allowed for the comparison and validation of self-reported data. We also obtained from Buoy the symptoms, diagnoses, triage, and sex of eligible users who opted not to participate in our study. This allowed us to compare our sample to the population of users and assess potential bias. In addition, a benefit of a collaborative approach is the potential to overcome the self-report limitations of prior studies. Thus, we obtained from Buoy the paths that individuals took and Buoy's final recommendation. We were also able to match the initial reason for the consultation to the reason reported in the survey and assess the extent to which respondents understood the recommendation and intended to act upon it. By leveraging a public or private partnership, we were able to explore the use and effects of a web-based symptom checker, which has important implications for health equity and the health care system during and after the COVID-19 crisis.

The limitations of this study include the use of cross-sectional data, which limited the ability to make any causal inferences, and the potential lack of applicability to other web-based symptom checkers. In addition, we did not assess the actual search terms entered by users. Finally, our study used a limited definition of web-based health information. Searches for symptoms using a web-based symptom checker differ from other forms of health-related information communicated through the internet. For example, web-based health communities can also be a source of social support [54] and peer-to-peer medical advice [7].

Conclusions

The results of this study demonstrate the potential utility of a web-based health information tool to empower people to seek appropriate care and reduce health-related anxiety. An interactive symptom checker might provide more personalized and potentially reliable medical information than other forms of web-based health information–seeking. Despite encouraging results suggesting that the web-based tool may fulfill unmet health information needs among women and Black and Latino adults, analyses of the user base illustrate persistent second-level digital divide effects.

For web-based symptom checkers to make a meaningful contribution, they must not only be trusted by users but also meet their diverse needs, especially those concerning usability and comprehensibility. The inability to access web-based symptom checkers may also be associated with increased disparities in access to care, particularly among groups that have lagged historically in terms of digital access and literacy. Moreover, web-based symptom checker business models may further exacerbate these disparities. In contrast, AI technologies such as Buoy have the potential to alleviate disparities by allowing users to access accurate, actionable, and personalized advice within an evolving but often confusing web-based health information environment. Finally, there is a lack of evidence on whether web-based symptom checkers influence care-seeking behaviors. To address this gap, future research will use Buoy users' follow-up data to assess the extent to which users discuss their web-based findings with physicians, as well as barriers to the same and patient satisfaction.

Acknowledgments

This project was supported by a gift from Buoy Health to the University of California, Merced. Buoy Health staff provided access to participants and internal data used to validate self-reported data but had no role in data analysis or interpretation. The authors gratefully acknowledge Jovo Velasco, Stephanie Reyes, and Michael Vang for their data collection and coding support in this research. The contents are solely the responsibility of the authors and do not necessarily represent the official views of Buoy Health or the University of California.

Conflicts of Interest

The authors received funding and some technical assistance in recruiting from Buoy Health, although they were not involved in the research design, data analysis, or interpretation.

Multimedia Appendix 1

Buoy triage levels and recommendations.
[DOC File , 38 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Symptom and diagnosis codes and frequencies.
[DOCX File , 24 KB-Multimedia Appendix 2]

References

1. 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]
2. McCarthy M. Sixty seconds on . . . symptom checkers. BMJ 2016 Oct 20;355:i5596. [doi: 10.1136/bmj.i5596] [Medline: 27765782]
3. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
4. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
5. Morita T, Rahman A, Hasegawa T, Ozaki A, Tanimoto T. The potential possibility of symptom checker. Int J Health Policy Manag 2017 Oct 01;6(10):615-616 [FREE Full text] [doi: 10.15171/ijhpm.2017.41] [Medline: 28949479]
6. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
7. Fox S. After Dr Google: peer-to-peer health care. Pediatrics 2013 Jun;131 Suppl 4:S224-S225. [doi: 10.1542/peds.2012-3786K] [Medline: 23729765]
8. Hartzband P, Groopman J. Untangling the Web--patients, doctors, and the Internet. N Engl J Med 2010 Mar 25;362(12):1063-1066. [doi: 10.1056/NEJMp0911938] [Medline: 20335581]
9. 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]
10. Tsai CH, You Y, Gui X, Kou Y, Carroll JM. Exploring and promoting diagnostic transparency and explainability in online symptom checkers. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-17. [doi: 10.1145/3411764.3445101]
11. Meyer AN, Giardina TD, Spitzmueller C, Shahid U, Scott TM, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
12. Katz E, Blumler JG, Gurevitch M. Uses and gratifications research. Public Opin Q 1973;37(4):509-523. [doi: 10.1086/268109]
13. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
14. Massey PM. Where Do U.S. adults who do not use the Internet get health information? Examining digital health information disparities from 2008 to 2013. J Health Commun 2016;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]

15. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
16. Ramirez AS, Leyva B, Graff K, Nelson DE, Huerta E. Seeking information on behalf of others: an analysis of calls to a Spanish-language radio health program. Health Promot Pract 2015 Jul;16(4):501-509. [doi: 10.1177/1524839915574246] [Medline: 25716191]
17. Abrahamson JA, Fisher KE, Turner AG, Durrance JC, Turner TC. Lay information mediary behavior uncovered: exploring how nonprofessionals seek health information for themselves and others online. J Med Libr Assoc 2008 Oct;96(4):310-323 [FREE Full text] [doi: 10.3163/1536-5050.96.4.006] [Medline: 18974809]
18. Kelly B, Hornik R, Romantan A, Schwartz JS, Armstrong K, DeMichele A, et al. Cancer information scanning and seeking in the general population. J Health Commun 2010 Oct;15(7):734-753 [FREE Full text] [doi: 10.1080/10810730.2010.514029] [Medline: 21104503]
19. Niederdeppe J, Hornik RC, Kelly BJ, Frosch DL, Romantan A, Stevens RS, et al. Examining the dimensions of cancer-related information seeking and scanning behavior. Health Commun 2007;22(2):153-167. [doi: 10.1080/10410230701454189] [Medline: 17668995]
20. Rooks RN, Wiltshire JC, Elder K, BeLue R, Gary LC. Health information seeking and use outside of the medical encounter: is it associated with race and ethnicity? Soc Sci Med 2012 Jan;74(2):176-184. [doi: 10.1016/j.socscimed.2011.09.040] [Medline: 22154611]
21. Grasso KL, Bell RA. Understanding health information seeking: a test of the risk perception attitude framework. J Health Commun 2015;20(12):1406-1414. [doi: 10.1080/10810730.2015.1018634] [Medline: 26161622]
22. Shim M, Kelly B, Hornik R. Cancer information scanning and seeking behavior is associated with knowledge, lifestyle choices, and screening. J Health Commun 2006;11 Suppl 1:157-172. [doi: 10.1080/10810730600637475] [Medline: 16641081]
23. Fareed N, Swoboda CM, Jonnalagadda P, Walker DM, Huerta TR. Differences between races in health information seeking and trust over time: evidence from a cross-sectional, pooled analyses of HINTS data. Am J Health Promot 2021 Jan;35(1):84-92. [doi: 10.1177/0890117120934609] [Medline: 32588638]
24. Jiang S, Street RL. Pathway linking Internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Aug;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
25. Ramírez AS, Freres D, Martinez LS, Lewis N, Bourgoin A, Kelly BJ, et al. Information seeking from media and family/friends increases the likelihood of engaging in healthy lifestyle behaviors. J Health Commun 2013;18(5):527-542 [FREE Full text] [doi: 10.1080/10810730.2012.743632] [Medline: 23472825]
26. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
27. Moldovan-Johnson M, Tan AS, Hornik RC. Navigating the cancer information environment: the reciprocal relationship between patient-clinician information engagement and information seeking from nonmedical sources. Health Commun 2014;29(10):974-983 [FREE Full text] [doi: 10.1080/10410236.2013.822770] [Medline: 24359259]
28. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the Internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
29. Cocco AM, Zordan R, Taylor DM, Weiland TJ, Dilley SJ, Kant J, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. Med J Aust 2018 Oct 15;209(8):342-347. [doi: 10.5694/mja17.00889] [Medline: 30107763]
30. Buoy Health: Check Symptoms & Find the Right Care. URL: https://www.buoyhealth.com/ [accessed 2022-08-10]
31. Mack H. Digital health startup Buoy launches AI-powered, symptom-checking chatbot. MobiHealthNews. 2017 Mar 8. URL: https://www.mobihealthnews.com/content/digital-health-startup-buoy-launches-ai-powered-symptom-checking-chatbot [accessed 2022-08-10]
32. Buoy Health - A chatbot that helps diagnose your symptoms. Product Hunt. 2017. URL: https://www.producthunt.com/posts/buoy-health [accessed 2022-08-10]
33. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of use of online symptom checkers with patients' plans for seeking care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]
34. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. Med Care 2005 Jun;43(6):607-615. [doi: 10.1097/01.mlr.0000163658.65008.ec] [Medline: 15908856]
35. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York, NY, USA: Psychology Press; 2010.
36. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-484. [doi: 10.1080/10810730490504233] [Medline: 15513791]

37. 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]
38. Sivakumar G, Mares ML. The doctor versus the Internet: effects of low-, medium-, and high-quality websites on intentions to follow the doctor's advice. Health Commun 2017 Dec;32(12):1454-1462. [doi: 10.1080/10410236.2016.1228030] [Medline: 27767346]
39. van Weert JC, van Noort G, Bol N, van Dijk L, Tates K, Jansen J. Tailored information for cancer patients on the Internet: effects of visual cues and language complexity on information recall and satisfaction. Patient Educ Couns 2011 Sep;84(3):368-378. [doi: 10.1016/j.pec.2011.04.006] [Medline: 21550757]
40. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
41. White RW, Horvitz E. Experiences with Web search on medical concerns and self diagnosis. AMIA Annu Symp Proc 2009 Nov 14;2009:696-700 [FREE Full text] [Medline: 20351943]
42. Schneider D, Appleton L, McLemore T. A reason for visit classification for ambulatory care. Vital Health Stat 2 1979 Feb(78):i-63 [FREE Full text] [Medline: 433160]
43. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
44. Shapiro MF, Ware Jr JE, Sherbourne CD. Effects of cost sharing on seeking care for serious and minor symptoms. Results of a randomized controlled trial. Ann Intern Med 1986 Feb;104(2):246-251. [doi: 10.7326/0003-4819-104-2-246] [Medline: 3946953]
45. Li J, Theng YL, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2016 Dec;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
46. Kwon JH, Kye SY, Park EY, Oh KH, Park K. What predicts the trust of online health information? Epidemiol Health 2015 Jun 28;37:e2015030 [FREE Full text] [doi: 10.4178/epih/e2015030] [Medline: 26212505]
47. 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]
48. Thai CL, Gaysynsky A, Falisi A, Chou WY, Blake K, Hesse BW. Trust in health information sources and channels, then and now: evidence from the health information national trends survey (2005–2013). In: Hale TM, Chou WY, Cotten SR, Khilani A, editors. eHealth: Current Evidence, Promises, Perils and Future Directions (Studies in Media and Communications, Vol. 15). Bingley, UK: Emerald Publishing Limited; 2018:43-67.
49. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
50. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
51. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
52. Kravitz RL, Callahan EJ, Paterniti D, Antonius D, Dunham M, Lewis CE. Prevalence and sources of patients' unmet expectations for care. Ann Intern Med 1996 Nov 01;125(9):730-737. [doi: 10.7326/0003-4819-125-9-199611010-00004] [Medline: 8929006]
53. Musa D, Schulz R, Harris R, Silverman M, Thomas SB. Trust in the health care system and the use of preventive health services by older black and white adults. Am J Public Health 2009 Jul;99(7):1293-1299. [doi: 10.2105/AJPH.2007.123927] [Medline: 18923129]
54. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]

Abbreviations

AI: artificial intelligence
NAMCS: National Ambulatory Medical Care Survey
OR: odds ratio
RQ: research question

Edited by T Leung; submitted 17.01.22; peer-reviewed by N Fareed, S Song, Y Qian; comments to author 16.03.22; revised version received 21.04.22; accepted 24.05.22; published 19.08.22

Please cite as:
Arellano Carmona K, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS
Health Information Seeking From an Intelligent Web-Based Symptom Checker: Cross-sectional Questionnaire Study
J Med Internet Res 2022;24(8):e36322
URL: https://www.jmir.org/2022/8/e36322
doi: 10.2196/36322
PMID:

Original Paper

Identifying Influences in Patient Decision-making Processes in Online Health Communities: Data Science Approach

Mingda Li[1], PhD; Jinhe Shi[1], PhD; Yi Chen[2], PhD

[1]Ying Wu College of Computing, New Jersey Institute of Technology, Newark, NJ, United States
[2]Martin Tuchman School of Management, New Jersey Institute of Technology, Newark, NJ, United States

Corresponding Author:
Yi Chen, PhD
Martin Tuchman School of Management
New Jersey Institute of Technology
184-198 Central Ave
Newark, NJ, 07102
United States
Phone: 1 973 596 6302
Email: yi.chen@njit.edu

Abstract

Background: In recent years, an increasing number of users have joined online health communities (OHCs) to obtain information and seek support. Patients often look for information and suggestions to support their health care decision-making. It is important to understand patient decision-making processes and identify the influences that patients receive from OHCs.

Objective: We aimed to identify the posts in discussion threads that have influence on users who seek help in their decision-making.

Methods: We proposed a definition of influence relationship of posts in discussion threads. We then developed a framework and a deep learning model for identifying influence relationships. We leveraged the state-of-the-art text relevance measurement methods to generate sparse feature vectors to present text relevance. We modeled the probability of question and action presence in a post as dense features. We then used deep learning techniques to combine the sparse and dense features to learn the influence relationships.

Results: We evaluated the proposed techniques on discussion threads from a popular cancer survivor OHC. The empirical evaluation demonstrated the effectiveness of our approach.

Conclusions: It is feasible to identify influence relationships in OHCs. Using the proposed techniques, a significant number of discussions on an OHC were identified to have had influence. Such discussions are more likely to affect user decision-making processes and engage users' participation in OHCs. Studies on those discussions can help improve information quality, user engagement, and user experience.

(J Med Internet Res 2022;24(8):e30634) doi: 10.2196/30634

KEYWORDS

influence relationship; decision-making threads; online health communities; patient engagement; deep learning; text relevance measurement

Introduction

Background

In recent years, online health communities (OHCs) such as the Cancer Survivors Network (CSN), MedHelp, DoctorLounge, WebMD, and Health-boards message boards have become one of the most important resources that patients leverage [1]. An OHC is defined as an asynchronous web-based message board system for patients that contains multiple message boards, each of which typically focuses on 1 disease. OHCs provide a web-based channel that enables information exchange, facilitates communication, and provides support to patients and caregivers [2-4]. They are especially valuable for patients with chronic diseases to learn about their conditions and seek social support [5,6].

Empowering and supporting patients to make informed health care decisions is a key component of patient-centered health care and is a social, economic, and technical necessity [7,8]. A lot of patients seek information and advice on OHCs. Existing work has found that nearly half of the threads in a breast cancer forum [9] are related to patient decision-making [1]. Studies have also shown that patients are often influenced by web-based sources and social media in their health care decision-making [10,11].

Objectives

The goal of this study was to identify the influence relationship of posts in discussion threads related to health care decision-making. Specifically, we defined the influence relationships and identified post replies that influenced the initial author, who had questions posted on OHCs.

The outcomes of this study are important for health care professionals to help patients make informed decisions for several reasons. First, analyzing the writing style and pattern of posts that have influence may help explain why they have influence and provide insights to health care professionals on effective communication with patients. Second, if the information provided by posts that have an influence is not accurate, it will mislead patients. It is important to check the information quality in such posts to improve the quality of influence. Furthermore, a patient who has questions but does not receive any replies that have an influence may need further help.

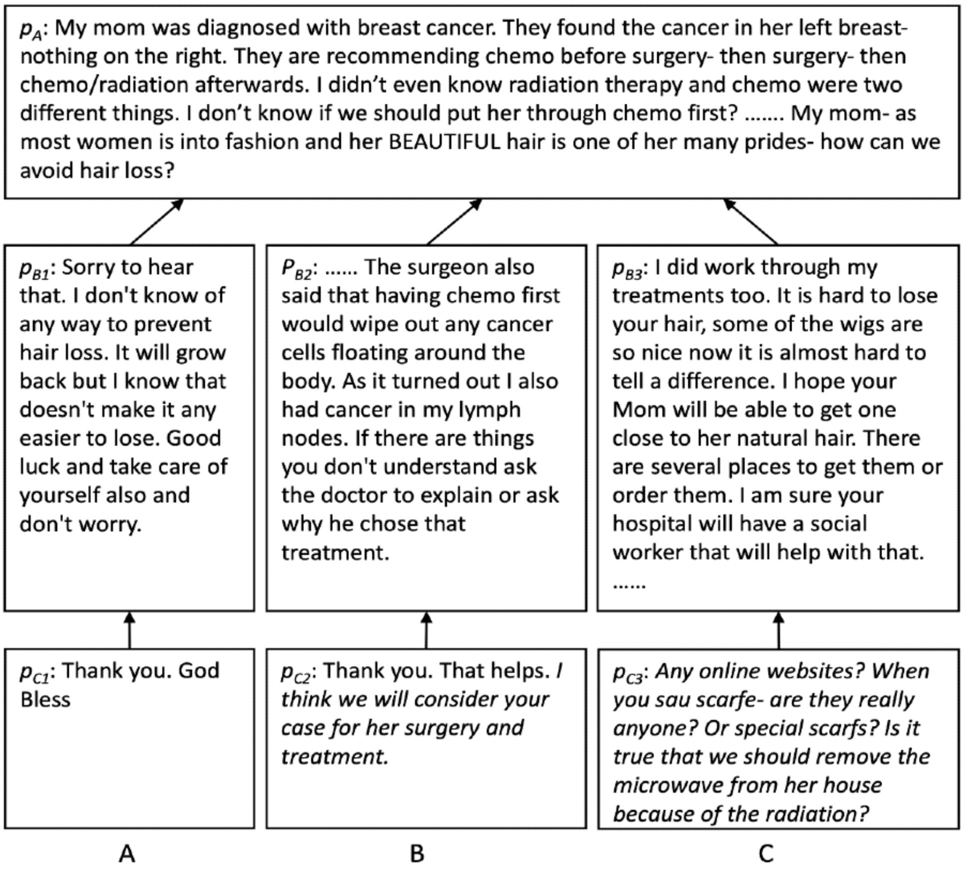
Literature Review

There is a lot of research conducted on OHC analysis, although with limited study on identifying influence relationships of posts. Several studies have been conducted on analyzing the reciprocal patterns between users' replies in discussion forums [12-14]. There is also work on analyzing the patterns between post views and post replies [15]. Many studies have been conducted on identifying influential users in a community [16-20]. In those applications, a post, blog, or tweet typically expresses an opinion of the author, and the replies are considered as an indication of being influenced by the opinion of the original post. That is, the reply relationship is considered as an influence relationship. The focus is on judging the influential power of an author based on activeness of post writing [21] and social network features [17,18] such as PageRank-like algorithms or clustering algorithms.

Finding influence relationships among posts in discussion forums is different from finding influential users and requires different techniques. In an OHC, the initial author of a thread typically expresses a question, not an opinion. The influence happens when a reply to the question affects the initial author. There are only 2 existing studies that consider the influence of the replier on the initial author [21,22]. This influence is identified when the sentiment of the initial author is changed to be similar to that of the replier. However, this definition may not be accurate.

Let us look at an example of a discussion thread related to patient decision-making, shown in Figure 1. An OHC user initialized a thread asking for advice on whether to have chemotherapy before surgery for her mother's treatment plan in post p_A.

Figure 1. Example of a discussion thread.



p_A: My mom was diagnosed with breast cancer. They found the cancer in her left breast- nothing on the right. They are recommending chemo before surgery- then surgery- then chemo/radiation afterwards. I didn't even know radiation therapy and chemo were two different things. I don't know if we should put her through chemo first? ……. My mom- as most women is into fashion and her BEAUTIFUL hair is one of her many prides- how can we avoid hair loss?

p_{B1}: Sorry to hear that. I don't know of any way to prevent hair loss. It will grow back but I know that doesn't make it any easier to lose. Good luck and take care of yourself also and don't worry.

P_{B2}: …… The surgeon also said that having chemo first would wipe out any cancer cells floating around the body. As it turned out I also had cancer in my lymph nodes. If there are things you don't understand ask the doctor to explain or ask why he chose that treatment.

p_{B3}: I did work through my treatments too. It is hard to lose your hair, some of the wigs are so nice now it is almost hard to tell a difference. I hope your Mom will be able to get one close to her natural hair. There are several places to get them or order them. I am sure your hospital will have a social worker that will help with that. ……

p_{C1}: Thank you. God Bless

p_{C2}: *Thank you. That helps. I think we will consider your case for her surgery and treatment.*

p_{C3}: *Any online websites? When you sau scarfe- are they really anyone? Or special scarfs? Is it true that we should remove the microwave from her house because of the radiation?*

In Figure 1A, a user replied by comforting her in post p_{B1}. The reply was not informative. Even though the initial author expressed gratefulness to the author of post p_{B1}, with sentiment changing to be positive in post p_{C1}, she was not influenced by post p_{B1}. Indeed, studies show that 75% to 85% of CSN forum participants change their sentiment in a positive direction through web-based interactions with other community members [23]. A change in sentiment is not necessarily an indicator of being influenced.

In contrast, in Figure 1B, a user shared her experience in a similar situation suggesting to have chemotherapy before a surgery in post p_{B2}. The initial author expressed her gratitude and indicated that she would consider this suggestion in determining her mother's treatment plan (the sentences in italics) in p_{C2}, showing her being influenced.

Contribution

Instead of considering sentiment changes, we propose using questions or future actions on relevant replies as an indicator of being influenced, as illustrated in the aforementioned example. There are 2 major challenges in identifying influence relationships. First, we need to define influence relationships of posts. We examined the semantics of post content to define influence relationships. Unlike influential users, who are defined by network features in the existing work [16-20], text content is the key to determine whether posts have influence. Second, it is hard to identify influence relationships. Unlike typical text classification problems, influence relationships involve multiple posts with reply relationships rather than a single paragraph of text. In addition, influence is an abstract concept. It is challenging to extract relevant features to capture the influence patterns considering both content and the reply relationship.

This study makes novel contributions to identifying influence relationships in discussion threads in OHCs related to patient decision-making. Specifically, (1) we defined the influence relationship between the posts based on the semantics of the post content, (2) an extensible deep learning model that extracts and combines both sparse and dense features was proposed to identify the influence relationships in OHC decision-making threads, and (3) the proposed model achieved good performance in identifying influence relationships in empirical evaluation.

Methods

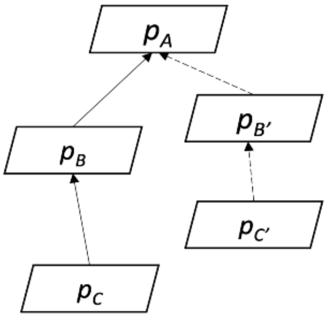
In this section, we first model the OHC data and define the influence relationship in discussion threads. We then propose a deep learning–based model to identify the influence relationships.

Problem Definition

Definition of Discussion Threads

Figure 2 presents an overview of the OHC data structure. We modeled an OHC as a set of discussion threads $T = \{t_1, t_2,..., t_n\}$. Each thread t_i is composed of a set of posts and a function R that represents the reply relationship. For example, Figure 2 illustrates a thread that contains a set of 5 posts $\{p_A, p_B, p_C, p_{B'}, p_{C'}\}$. One of the reply relationships, $R(p_B) = p_A$, represents that post p_B replies to post p_A. Each post p_i consists of a sequence of sentences $p_i = \{s_1, s_2,..., s_l\}$. Each post has an author. We denoted the author relationship using a function U. $U(p_i)$ represents the author of post p_i. Note that a post only has a single author; however, an author may write ≥0 posts in a thread. We used p_A to present the first post of a thread and named it the *initial post*. The author of the *initial post*, $U(p_A)$, is referred to as the *initial author* of the thread.

Figure 2. Data structure of an online health community.



Existing work [1] has studied the thread discussions in OHCs and identified that a subset of threads is related to *patient decision-making*. Such a thread is characterized by questions in the initial post and replies with suggestions of options. Techniques have been developed to identify decision-making threads in OHCs.

In this paper, we study how to identify the cases where the initial author of a decision-making thread is influenced by a reply post. Note that our study is general to any thread discussions related to decision-making. The definition and identification of decision-making threads can be handled using the approach developed in existing work [1] or other approaches. In the rest of this paper, we use *threads* to refer to decision-making threads for simplicity. The defined influence relationship may not be applicable to discussion threads that are not related to decision-making, such as discussion threads for casual communication or experience-sharing threads providing social support.

Definition of Relationships

Overview

Before introducing the definition of *influence relationships*, we first introduce relationships. A relationship is defined on a triple of posts in a thread with reply relationships: an initial post, a reply to the initial post, and the initial author's subsequent reply.

Definition 1 (Relationship)

We define the relationship among three posts p_A, p_B, and p_C, in a thread as $r_i = (p_A, p_B, \text{and } p_C)$, where post p_A is the initial post of the thread, post p_B replies to p_A, post p_C replies to p_B, and the authors of p_A and p_C are the same person. That is, $R(p_B) = p_A$, $R(p_C) = p_B$, and $U(p_A) = U(p_C)$.

We used $r_i = (p_A, p_B, p_C)$ to denote the relationship among p_A, p_B, and p_C. Note that there are many such relationships in a thread, and we considered all such triples. For instance, Figure 2 shows a thread with 2 relationships, $r_1 = (p_A, p_B, p_C)$ and $r_2 = (p_A, p_{B'}, p_{C'})$.

Also, note that existing work on identifying influential users [16-20] does not consider the relationships among post triples but only considers the reply relationship between 2 posts.

Definition of Influence Relationships

Intuition

Now, we discuss how to define *influence relationships* on relationship (p_A, p_B, p_C), where post p_B has an influence on the initial author $U(p_A)$.

First, intuitively, if post p_B influences the initial author $U(p_A)$, then the content of these 3 posts must be relevant.

Second, we referred to the definition of *influence* in Merriam-Webster [24]—"to affect or alter by indirect or intangible means"—and the reaction of *being influenced* is to *sway* rather than being convinced. If the initial author considers the suggestion given in post p_B, even if she eventually does not take the suggestion, she is considered to have been influenced by post p_B. On the basis of this definition, we observed 2 indications that the initial author, $U(p_A)$, was influenced by p_B.

An observation of being influenced is that the initial author may ask questions in p_C based on the suggestions in p_B. Curiosity is a motivator for learning and influential in decision-making [25]. An existing study [26] used a statistically large sample of learning forum posts to investigate whether student participation in the forum could be influenced. They observed that students who were influenced by others' interesting answers were more likely to ask follow-up questions. This indicates that asking further questions is a sign of being influenced. The same pattern also exists in OHCs. Let us look at the example in Figure 1C. The initial author expressed concerns about hair loss in p_A. Another user replied in post p_{B3} suggesting the use of wigs. The initial author then replied in post p_{C3} with questions (the sentences in italics) for more details about the suggestion given in post p_{B3}. These questions indicate that the initial author was thinking about the suggestion given in post p_B; that is, being influenced.

The second indication that the initial author was influenced by a post p_B is that she expressed her intention to take action in post p_C. Adjei et al [27] found that member-to-member communication in web-based brand communities greatly influenced the members' future purchase behavior. Similarly, the communication through discussion threads in OHCs may also affect the initial author's future actions. Let us look at the example in Figure 1B again. For the treatment question asked in p_A, a forum user shared her experience and discussed the treatment in post p_{B2}. The initial author then replied with a planned action (the sentence in italics) in p_{C2}. The intention of future action based on the communications in the thread is an indicator of the influence relationship.

On the basis of these observations, we define influence relationships in decision-making threads in the following section.

Definition 2 (Influence Relationship)

A relationship $r_i = (p_A, p_B, p_C)$ is considered as an influence relationship—that is, $U(p_A)$ is influenced by p_B—if and only if the following conditions are met: (1) the content of p_B is relevant to post p_A, (2) the content of p_C is relevant to post p_B, and (3) p_C contains questions or indicates future actions.

To identify influence relationships, we modeled it as a classification task. Given a set of relationships $R = \{r_1, r_2,..., r_n\}$, for each relationship r_i, we predicted its label to be either 1 or −1, where label 1 indicated that r_i was an influence relationship and label –1 indicated that r_i was not an influence relationship. The goal was to learn a model from the labels of known relationships and predict the labels for unlabeled relationships.

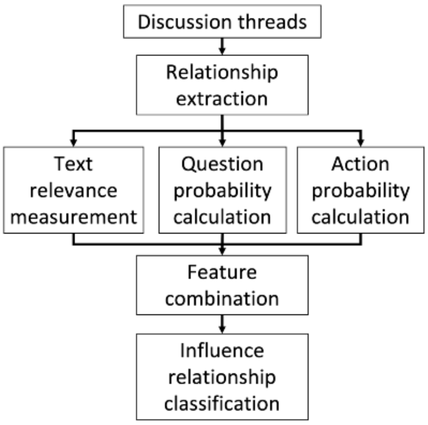
Model Design

Overview

In this section, we present the method to identify the influence relationships in decision-making threads in OHCs. Figure 3 presents the framework of the proposed method.

Given a set of discussion threads as the input, we first extracted the triple relationships using the relationship extraction module. Text relevance features, question probability features, and action probability features were then calculated using the text relevance measurement module, the question probability calculation module, and the action probability calculation module, respectively. Finally, all the features were combined using a deep learning model in the feature combination module to generate the probability of a relationship being an influence relationship.

Figure 3. Workflow of influence relationship identification.



Relationship Extraction Module

In this section, we introduce the relationship extraction module, which extracted all relationships defined in definition 1.

In the first step of relationship extraction, we built the reply tree structure based on the indented format in html files. For each adjacent post pair, the post that was posted earlier was treated as the parent of the latter post. The ancestor-descent distance between a post and the initial post was represented by the number of tab characters. The reply structure of a thread is illustrated in Figure 2. Each post is a node in the thread tree, and each edge represents a reply relationship. The root of the thread tree is the initial post (ie, p_A) in definition 1.

Existing work observes that, in some forums, the reply structure in a discussion thread may not be fully available and proposes techniques to construct full reply structures [28]. The OHCs used in our experiments had a full reply structure. Existing techniques can be leveraged if needed for other forums.

We then navigated the thread tree to extract all relationship triples, as defined in definition 1. Each triple started with the initial post followed by a reply to the initial post written by another author and then a subsequent reply by the initial author, all of which were on the same path in the thread tree. For example, $r_1 = (p_A, p_B, p_C)$ and $r_2 = (p_A, p_{B'}, p_{C'})$ are 2 relationships in the thread tree in Figure 2.

Text Relevance Measurement Module

The text relevance measurement module measures the content relevance, or text semantic similarity, of 2 posts using a relevance score between 0 and 1.

There are mainly 2 types of deep learning–based methods in the literature that measure text relevance. The first type of method extracts content feature vectors of 2 input texts and then combines them to make a prediction, such as the Deep Structured Semantic Models (DSSM) [29], the Convolutional DSSM [30], and Architecture-I (ARC-I) [31]. The intuition of this method is to highlight the important information of the original texts so that irrelevant content can be removed before the feature combination phase. However, the drawback of this type of method is that it runs the risk of losing detail [32].

The second type generates the word-level relevance first and then uses neural networks to learn the hierarchical interaction patterns for content-level relevance, such as DeepMatch [33], Architecture-II (ARC-II) [31], and MatchPyramid [34]. The motivation is that making a good relevance judgment requires considering the interactions in the text relevance measurement process, starting from the interactions between words to patterns in phrases and those in whole sentences [34]. However, the training process for the second type is much more expensive than for the first one.

We evaluated both approaches to measure text relevance in experiments. We chose 2 state-of-the-art representative methods for the text relevance measurement module in the evaluation. For the first type, we chose ARC-I [31], which uses a multilayer perceptron to combine relevance feature vectors. It shows better performance than the DSSM [29] and Convolutional DSSM [30], both of which use cosine similarity [34]. We chose MatchPyramid [34] to represent the second type of method as it exhibits better performance than the other 2 methods (DeepMatch [33] and ARC-II [31]) in experiments on multiple data sets [34].

We further proposed the adaptation of Bidirectional Encoder Representations from Transformers (BERT) [35] as the embedding layer in the ARC-I and MatchPyramid models. BERT is a state-of-the-art embedding method for word representation in many natural language understanding tasks, trained on BookCorpus and English Wikipedia. We considered both BERT (trained on Wikipedia) and word2vec (trained on the training data set) as the embedding methods for both ARC-I and MatchPyramid. Different variations of the text relevance measurement module are evaluated in the *Text Relevance Evaluation* section.

Question Probability Calculation Module

We now discuss how to calculate the probability of a post containing a question using the question probability calculation module.

There are 2 types of methods to identify question sentences in forums: a rule-based approach and a learning-based approach. In a rule-based approach, question marks and 5W1H words (what, who, when, where, why, and how) are used to identify question sentences [36]. A learning-based approach uses sequential question patterns to train a binary classifier on labeled data [37-40]. Liu and Jansen [37] used the question mark to extract question posts from Sina Weibo. In the studies by Ranganath et al [38,39], frameworks were proposed to identify rhetorical questions by modeling the motivation of the user for posting them. In the study by Ojokoh et al [40], questions from ResearchGate were identified based on the maximum probability value of a naïve Bayes classification with part-of-speech tag features.

Both rule-based and learning-based approaches can achieve excellent performances. A study shows that a rule-based approach can outperform complicated learning-based approaches [36]. Thus, we followed a rule-based method [36] to identify question presence in the posts. In total, 2 types of rules were considered: question marks and 5W1H words. We made adaptations of this approach for OHCs. As a question mark is the most significant sign of a question, we gave a higher confidence score to a sentence with a question mark. We also set some constraints on 5W1H words to simulate the pattern of question sentences. First, 5W1H must appear at the beginning of a sentence. Second, auxiliary words were added to the original words for more specific patterns. For example, we considered *what is*, *what are*, *what does*, and *what do* instead of *what*.

After the question probability of each sentence in a post p_i was calculated, the maximum probability was used as the likelihood of post p_i containing at least one question, denoted as $Q(p_i)$.

Action Probability Calculation Module

This section presents the action probability calculation module, which generated the probability of action presence in a post.

The indication of a future action can be captured by the presence of verbs and appropriate sentence tense. The Natural Language Toolkit (NLTK) [41] tagger module defines a standard interface for augmenting each token of a text with supplementary information, such as its part of speech or its WordNet synset tag, and provides several different implementations for this interface. We leveraged the NLTK tagger module to assess the likelihood of a post containing future actions by checking the existence of words with a future tense verb tag (eg, *will consider* in Figure 1B) or a modal auxiliaries tag (eg, *can*, *could*, *may*, and *must*). To count on the cases where future tenses may not be identified because of forum users' typos or informal writing, we set the probability of future action to be 0.5 when the rules failed to identify future actions. Equation 1 shows the calculation formula to generate the action probability of a post p_i.

$$A(p_i) = \begin{cases} 1 & if\ VBF\ or\ MD \in POS(p_i) \\ 0.5 & otherwise \end{cases} \quad \textbf{(1)}$$

Note that we did not consider negation in the action probability calculation module. For example, in post p_C, the initial author disagrees with the suggestions proposed in p_B and decides to do something different. For those cases, the overall meaning of p_B and p_C would be the opposite and, therefore, would be captured by the relevance vectors generated in the text relevance measurement module. Thus, we did not consider negations in this phase to avoid double counting.

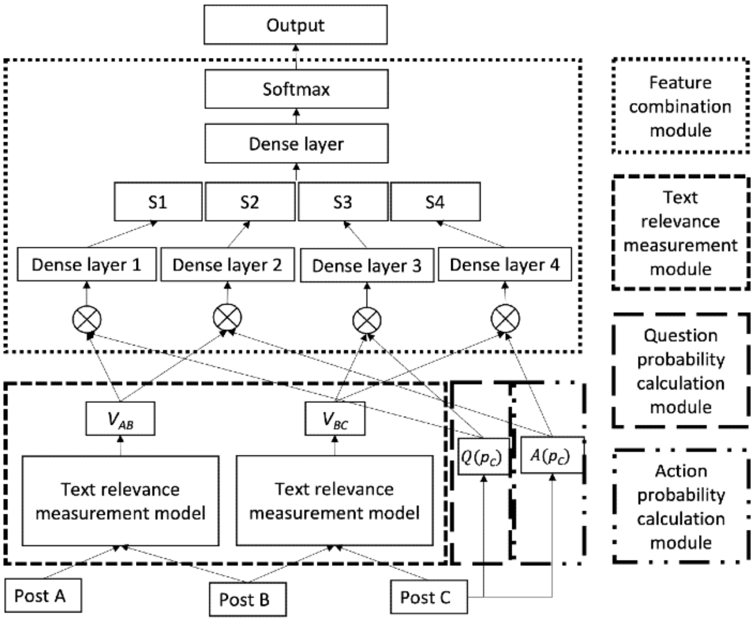
Feature Combination Module

Overview

Referring to Figure 4, the text relevance measurement module calculated P_{AB}—the relevance score between p_A and p_B—and P_{BC}—the relevance score between p_B and p_C. The question probability calculation module and action probability calculation module calculated the question probability $Q(p_C)$—or Q in short—and action probability $A(p_C)$—or A in short—based on the text of p_C.

We now discuss the feature combination module that measures the influence score based on these features. We discuss 2 alternative methods: *a baseline approach* and *a deep learning model*.

Figure 4. Architecture of the feature combination module.



Baseline Approach

Recall that, according to definition 2, the presence of an influence relationship requires the relevance between post p_A and post p_B, the relevance between post p_B and post p_C, and the presence of a question or action in post p_C. We started with an intuitive method to detect influence relationships based on the definition using Equation 2.

$$P_{baseline} = P_{AB} \times P_{BC} \times \max\,[Q(p_C), A(p_C)] \quad \textbf{(2)}$$

We set the thresholds to 0.5, 0.5, and 0.9 for each component.

Deep Learning Approach

We further proposed a deep learning model that combines the text relevance, the likelihood of question presence, and the likelihood of future action presence to identify influence relationships. The architecture of this model is shown in Figure 4.

Compared with the *baseline* approach, there are 3 major benefits of using a deep learning model. First, it is labor-intensive, time-consuming, and difficult to determine appropriate thresholds for cutting off the probabilities using a rule-based approach such as the *baseline* approach. A threshold that works well for one data set may not be optimal for another. Both a rule-based approach and a deep learning model require different thresholds for different data sets. A rule-based approach requires manual parameter tuning for each data set. In contrast, a deep learning approach learns thresholds from the ground truth and, thus, can easily adapt to a new data set with minimal human intervention [42]. Second, the question and action features may have different interactions with the relevance features. We observed that questions are often relevant, but actions are not necessarily. People typically express appreciation in post p_C or sometimes even mention actions totally irrelevant to post p_B, such as the plan to travel or shop. Being relevant is more important to consider in the presence of actions compared with in the presence of questions. However, in the *baseline* approach, the question and action features are merged before being combined with the relevance features, resulting in the loss of important information. Furthermore, we used relevance vectors as inputs to the deep learning model to calculate the influence score. Compared with the *baseline* approach, which uses the relevance scores as input to measure the influence score, relevance vectors provide much richer information. This can be especially helpful when there are several topics involved in the discussion. The relevance information is also leveraged during the phase of combining the relevance features with the question or action features.

Let V_{AB} denote the relevance vector between p_A and p_B and V_{BC} denote the relevance vector between p_B and p_C. We generated V_{AB},V_{BC} from p_A, p_B, and p_C and calculated Q and A from p_C.

These features were then connected. The question or future action in p_C must be related to the content of p_A and p_B. Thus, we combined V_{AB} and V_{BC} with Q and A using one of the following two operators: (1) *cat* (concatenating each relevance vector with question or action probability) and (2) *dot* (multiplying each relevance vector with question or action probability).

There are 2 major differences between these 2 operators for connecting the features: *cat* and *dot*. First, *dot* makes sure that Q and A affect each dimension in the relevance vectors, whereas *cat* cannot guarantee this as some neurons or nodes are dropped out. Some interactions between questions or actions and text relevance may be ignored by the *cat* operator. Second, the training process of the *cat* is more expensive than that of the *dot* because, for each dense layer 1 to 4, there is an additional dimension for the *cat* compared with for the *dot*.

In Figure 4, we use ⊗ to present the combination operator, which can be either *cat* or *dot*. The combination step produces 4 feature vectors: $V_{AB} \otimes Q$, $V_{AB} \otimes A$, $V_{BC} \otimes Q$, and $V_{BC} \otimes A$. To extract the key information from these combined feature vectors, 4 dense (fully connected) layers were used to populate the summarized feature vectors (S_1, S_2, S_3, S_4). The concatenation of these 4 summarized feature vectors was passed through 2 dense layers. The first one was used to further combine the summarized feature vectors. The second one aimed to generate the probability distribution over the labels. To avoid gradient vanishing and exploding [43], we chose the *Relu* function as the activation function for all the dense layers except the output layer, which uses the *softmax* function to populate the probabilities.

We trained the model using the binary cross-entropy loss function defined in Equation 3, which minimizes the distance between the probability distributions of the ground truth and those of the predicted score.

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} y_i \cdot \log\left(s_i\right) + \left(1 - y_i\right) \cdot \log\left(1 - s_i\right) \qquad \textbf{(3)}$$

Where y_i is the ground truth label of the ith training sample and s_i is the score predicted by the model. The Adam optimizer [43] was leveraged for optimization because of its advantage of processing sparse features and obtaining faster convergence compared with the normal stochastic gradient descent with momentum.

Ethics Approval

All materials were obtained from anonymous open-source data. Thus, ethics approval was not required.

Results

Experiment Setting and Evaluation Metrics

We implemented a prototype system for influence relationship identification on discussion threads. The prototype system and data sets used in the evaluation are publicly available at GitHub [44].

For empirical evaluation, we collected 25,208 threads that were publicly available in the CSN breast cancer forum [9]. The webpages were collected and processed by a web crawler we developed leveraging the Spider Crawler library [45]. There were 321,000 posts with 1.9 million sentences in total. We applied the classifier proposed by Li et al [1] on all 25,208 threads to identify the ones that were related to patient decision-making and obtained 11,815 (46.87%) such threads. Note that other models for classifying decision-making threads can also be plugged in.

We then extracted relationships from the decision-making threads using the relationship extraction module and obtained 9053 relationships. We randomly picked 853 (9.42%) of them to label. A total of 4 PhD students worked on the manual labeling. All the relationship triples and post pairs were first independently labeled. In case of disagreement, a consensus was reached after discussion. A total of 261 relationships were labeled as influence relationships. Recall that, per definition 1, each relationship is presented as a triple (p_A, p_B, p_C). We also labeled whether posts p_A and p_B were relevant (ie, P_{AB}) and whether posts p_B and p_C were relevant (ie, P_{BC}). We observed some reply posts with content expressing only comfort or wishes. Although they express care about the initial author's conditions and seem relevant, they are generic. After discussion, we reached an agreement that, when the initial post and reply post shared similar medical terms (such as *chemotherapy* and *chemo*), we would label them as relevant. All 1706 post pairs (p_A, p_B) and (p_B, p_C) of the 853 relationships were labeled. Of the 1706 pairs, 1210 (70.93%) were relevant pairs, and the remaining 496 (29.07%) were irrelevant. We split the set of relationships into a training set (90%) and a testing set (10%). The post pairs in the aforementioned training and test sets were used for text relevance training and testing, respectively.

The metrics used for evaluation included precision, recall, F_1 score, accuracy, area under the receiver operating characteristic curve (ROC AUC), and area under the precision-recall curve (PR AUC). They evaluated the effectiveness of a system using different aspects: (1) *precision*, also known as positive predictive value, is the fraction of relevant instances among the retrieved instances; (2) *recall*, also known as sensitivity, is the fraction of relevant instances that are retrieved among all relevant instances; (3) *F_1 score* measures a model's performance by calculating the harmonic mean of the precision and recall, as shown in the following equation: $F_1 = 2 \times \frac{precision \times recall}{precision + recall}$ **(4)**; (4) *accuracy* is a common evaluation metric for binary classification problems and is defined as the fraction of corrected predictions among the total number of predictions; (5) *ROC AUC* is a common evaluation metric for binary classification problems and is created by plotting the true positive rate against the false positive rate at various threshold settings; and (6) *PR AUC* is commonly used to evaluate the performance of a model on data sets with imbalanced labels.

Text Relevance Evaluation

Table 1 presents the classification results of the text relevance measurement module. In total, 2 observations were made. The first observation was that the models using BERT achieved high recall but low precision, whereas the models with word-embedding vectors trained on OHC data obtained balanced precision and recall values. There are 2 reasons for these results. First, OHC data are domain-sensitive and can benefit from domain-specific word representation. Second, the BERT transformer tends to link words in adjacent sentences by mistake. In the text relevance measurement module, precision was more important than recall as the accuracy of influence relationship identification depended on the precision of relevance classification. Thus, we used the word vectors trained on OHC data instead of BERT in the following experiments.

Table 1. Text relevance measurement module results.

	Precision	Recall	F_1	Accuracy	ROC AUC[a]	PR AUC[b]
MatchPyramid with BERT[c] (trained on Wikipedia)	0.578	*0.992* [d]	0.730	0.512	0.502	0.583
MatchPyramid with word2vec (trained on the training data set)	0.781	0.820[d]	*0.806*	0.692	0.763	0.854
ARC-I[e] with BERT (trained on Wikipedia)	0.523	0.890[d]	0.659	0.503	0.493	0.554
ARC-I with word2vec (trained on the training data set)	*0.832*	0.747[d]	0.785	*0.784*	*0.848*	*0.903*

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] PR AUC: area under the precision-recall curve.

[c] BERT: Bidirectional Encoder Representations from Transformers.

[d] The *P* value is statistically significant at *P*=.05.

[e] ARC-I: Architecture-I.

The second observation was that, with word vector embedding, ARC-I achieved a better performance than MatchPyramid in most of the evaluation metrics. In the ARC-I model, each input text goes through an embedding layer, a convolution layer, and a max pooling layer, and the extracted feature vectors are then concatenated together as the input to a fully connected layer that calculates the predicted relevance scores. MatchPyramid populates the local word relevance matrix first. Each cell of the matrix presents the dot product of the word-embedding vectors of the words in the text input. The patterns of these interactions are then extracted using a convolutional neural network [46]. Thus, ARC-I focuses on checking relevance based on the meaning of the whole text, whereas MatchPyramid focuses on summarizing the important relevance features based on local word similarity. For OHC data sets, posts were relatively long and often contained noisy information; thus, considering the meaning of the entire post text was more important than focusing on adjacent words. This is why the performance of ARC-I was better than that of MatchPyramid in our evaluation. We also observed that *ARC-I* with word2vec outperformed

MatchPyramidwith word2vec in both *ROC AUC* and *PR AUC* but had an inferior F_1 *score*. Note that F_1 averages the performance of all the samples by combining the precision and recall, whereas the ROC AUC and PR AUC cumulate the precisions among all samples with different recall thresholds. This indicates that the average performance of *MatchPyramidwith word2vec* was better, but the overall performance of *ARC-Iwith word2vec* was better.

Question and Action Probability Evaluation

Now, we present the evaluation of the question probability calculation module and the action probability calculation module. The performance is shown in Table 2. Good performance was achieved for question identification. For future action identification, a high score was achieved on recall but not on precision. The following are a few examples of posts that are classified as containing future actions but actually do not have action intent: *I will tell you though I hated my silicone* or *I would worry about it*. These sentences have verbs in the future tense, but those verbs only convey opinions or feelings rather than taking action on health care. We plan to improve action detection by training action sentence models as future work.

Recall that in the *deep learning* approach, question and action probabilities are considered as input features instead of imposing a strict requirement on their presence. We conducted an analysis on the test data in terms of their presence. All positive cases either had a probability of action presence of 1.0 or had a high probability of question presence, with an average probability of 0.986 (SD 0.033). This indicates that the *deep learning* approach captures definition 2 well, ensuring the high likelihood that either a question or a future action is present.

Table 2. Question and action calculation module results.

	Precision	Recall	F_1	Accuracy	ROC AUC[a]	PR AUC[b]
Question probability calculation module	*1.000*	*1.000* [c]	*1.000*	*1.000*	*1.000*	*1.000*
Action probability calculation module	0.771	*1.000* [c]	0.871	0.810	0.733	0.771

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] PR AUC: area under the precision-recall curve.

[c] The *P* value is statistically significant at *P*=.05.

Influence Relationship Classification Evaluation

Table 3 shows the performance of the *baseline* and *deep learning* approaches with alternative ways to combine text relevance vectors, question features, and action features. Recall that, for the feature combination module, *baseline* combines the text relevance score, the likelihood of question presence, and the likelihood of future action presence to identify influence relationships. *MatchPyramid+cat Q/A* represents the model using *MatchPyramid* to calculate the text relevance score and *cat* as the combination operator ⊗, whereas *MatchPyramid+dot Q/A* uses *dot* as the combination operator ⊗. *ARC-I+cat Q/A* represents the model using *ARC-I* to calculate the relevance score and *cat* as the combination operator ⊗, whereas *ARC-I+dot Q/A* uses *dot* as the combination operator ⊗.

Table 3. Influence relationship classification results.

	Precision	Recall	F_1	Accuracy	ROC AUC[a]	PR AUC[b]
Baseline	0.300	0.231[c]	0.261	0.595	0.495	0.307
MatchPyramid+cat Q/A[d]	0.667	0.154[c]	0.25	0.714	0.560	0.442
MatchPyramid+dot Q/A[e]	0.633	*0.577* [c]	*0.603*	0.667	0.634	0.481
ARC-I+cat Q/A[f]	0.667	0.154[c]	0.25	0.714	0.637	0.515
ARC-I+dot Q/A[g]	*0.750*	0.462[c]	0.571	*0.786*	*0.724*	*0.631*

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] PR AUC: area under the precision-recall curve.

[c] The *P* value is statistically significant at *P*=.05.

[d] MatchPyramid+cat Q/A: model using *MatchPyramid* to calculate the text relevance score and *cat* as the combination operator ⊗.

[e] MatchPyramid+dot Q/A: model using *MatchPyramid* to calculate the text relevance score and *dot* as the combination operator ⊗.

[f] ARC-I+cat Q/A: model using *Architecture-I* to calculate the relevance score and *cat* as the combination operator ⊗.

[g] ARC-I+dot Q/A: model using *Architecture-I* to calculate the relevance score and *dot* as the combination operator ⊗.

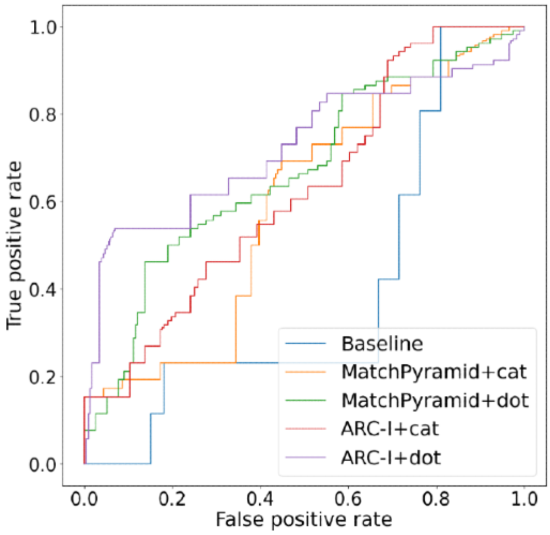
We also visualized the operating characteristic curves of all methods, as shown in Figure 5. From Table 3 and Figure 5, we have the following observations.

First, all proposed *deep learning* methods, which use relevance features and consider the interaction between relevance and the presence of questions or actions, significantly outperformed the *baseline* approach. This indicates that the relevance feature vectors generated by the text relevance measurement module were effective in capturing relevant content. Combining these feature vectors with the features of question presence and action presence helped capture their interactions and achieved good performance in influence relationship classification. In contrast, the *baseline* approach, which directly follows definition 2, did not perform well. This was due to the inability to capture the interactions between text relevance and question or action presence and the challenge of manually setting an appropriate cutoff threshold for each module.

Second, the models using the *dot* operator performed better than those using the *cat* operator. There are mainly 2 reasons for this. First, question probability and action probability may interact with V_{AB} and V_{BC} relevance vectors, which can be captured well by the *dot* operator. Figure 1B shows an example in which the action in p_C is related to the discussion in p_A and p_B. The action in p_{C2} is related to *chemo*, which is the common content of p_A and p_{B2}. In this case, the action probability needs to be combined with V_{AB}. Although, in another case, the action refers to an option mentioned in p_B, the interaction between p_B and p_C is more likely to be the context of the action and, thus, the action probability needs to be combined with V_{BC}. In contrast, the *cat* operator ignores some interactions between questions (actions) and the context because of the dropout of some neutrals. Therefore, the *cat*-based methods had a much lower recall than the *dot*-based methods. The results show that interactions between action and context are important for influence identification.

Furthermore, the *ARC-I+dot Q/A* had a much better precision, accuracy, ROC AUC, and PR AUC than *MatchPyramid+dot Q/A* but had lower recall and slightly lower F_1. This is because ARC-I achieved a better performance than MatchPyramid in the text relevance measurement module. *ARC-I+dot Q/A* was stricter than *MatchPyramid+dot Q/A* when fitting the model to the relevance factor. For applications that want to analyze the writing style and patterns of posts that have influence, precision is critical. *ARC-I+dot Q/A* is effective for locating such discussions. In contrast, for applications that want to check the information quality of the posts that have influence to prevent and mitigate the spread of misleading information, *MatchPyramid+dot Q/A* is more suitable because of its higher recall.

Figure 5. Influence relationship classification.



A Case Study

Figure 1 shows an example of 3 relationships, (p_A, p_{B1}, p_{C1}), (p_A, p_{B2}, p_{C2}), and (p_A, p_{B3}, p_{C3}), where (p_A is the initial post of the thread. The scores of these 3 relationships calculated using our system were 0.282, 0.793, and 0.622, respectively. Our system identified (p_A, p_{B2}, p_{C2}) and (p_A, p_{B3}, p_{C3}) as each containing an influence relationship, and (p_A, p_{B1}, p_{C1}) does not. As we can see from the post content, p_{B2} provides suggestions to the initial author regarding the treatment decision. In post p_{C2}, the initial author expresses actions to take based on the suggestions in p_{B2}. In post p_{B3}, the replier recommends that the author use wigs. The initial author then asks further questions about the wig information. Both relationships indicate that the initial author was influenced. In contrast, p_{B1} discusses general information and comforts the initial author, and the initial author expresses thanks in p_{C1}, but there is no indication of being influenced.

Discussion

Principal Findings

To the best of our knowledge, this is the first study that defines the influence relationships of discussion posts related to decision-making in OHCs. We proposed a deep learning–based natural language processing prototype to identify influence relationships. We then applied the developed techniques to identify the influence relationships in an OHC, the CSN breast cancer forum. There were 2 major observations.

First, we found that there is a significant amount of influence relationships in the OHC. Of the 9052 relationships in decision-making threads identified by Li et al [1], 3069 (33.9%)

were identified as influence relationships. That is, approximately one-third of the communications influence the initial authors on their decision-making. Furthermore, of the 5143 decision-making threads, which have at least one relationship, 2417 (47%) contain at least one influence relationship. Owing to the prevalence, it is important to study posts that have influence.

Second, we also observed that posts that have influence may contribute to engaging users in discussions. The average number of posts in threads containing at least one influence relationship was 15.5, whereas the average number of posts in threads containing no influence relationship was 12.6. Our conjecture is that posts that have an influence likely provide helpful information or good reasoning, which are thought-provoking and help engage users in discussions.

On the basis of these observations, there are several applications that can benefit from the identification and analysis of influence relationships.

First, analyzing the quality of posts that have influence helps improve the quality of the influence. As discussed in the first observation, influence relationships are common. Quality checking of those posts is more critical than that of other posts in terms of improving the effect of influences and mitigating the spread of misleading information.

On the basis of the identification of influence relationships, we can further identify influential users in OHCs. We can use existing techniques that analyze the network features to identify influential users [16-20], where this work calculates the edge weights (ie, the influence of a post). Identifying and checking influential users contributes to high-quality information dissemination.

Second, based on the second observation, analyzing the writing style of posts that have influence provides insights to health care professionals about effective communication for patient engagement.

Furthermore, identifying influence relationships contributes to effective information recommendations for addressing the information overload problem. When a user searches for information in OHCs, it is important to rank discussion threads and posts and recommend to users the most relevant and helpful discussions. On the basis of the analysis of influence relationships and the second observation, discussions that contain influence relationships are more likely to provide helpful information and encourage patient engagement. Thus, the presence of influence relationships is a positive factor in ranking.

Limitations

Our results are not without limitations. First, our definition of relationship was based on 3 posts, including the initial post in the thread. Therefore, we only identified the posts that had an influence on the initial author. However, any 3 posts that have a sequential reply relationship with the first and third posts from the same author can represent a relationship. We conjecture that the proposed techniques can be used to identify influence relationships among the generalized relationships and plan to study that problem in the future. Second, in this study, we considered text relevance between the posts in the relationship. Sometimes, even though 2 posts, p_B and p_C, are relevant overall, the specific sentence that has a question or future action indication in p_C may not be relevant to the suggestions in p_B. In addition, the current technique for future action detection sometimes generates false positives. To address these issues, we will investigate how to leverage part-of-speech and reference resolution techniques [47] to improve natural language understanding.

Conclusions and Future Work

We studied the problem of identifying influence relationships of web-based discussions and developed techniques and a prototype system for identifying influence relationships in OHCs. The proposed deep learning model demonstrates the performance advantage of the compared methods. As future work, we will address the aforementioned limitations to improve the generality and accuracy of the proposed techniques.

Acknowledgments

This material is based on work partially supported by the Leir Foundation and a grant from the National Institutes of Health (UL1TR003017).

Conflicts of Interest

None declared.

References

1. Li M, Shi J, Chen Y. Analyzing patient decision making in online health communities. In: Proceedings of the 2019 IEEE International Conference on Healthcare Informatics. 2019 Presented at: ICHI '19; June 10-13, 2019; Xi'an, China p. 1-8. [doi: 10.1109/ICHI.2019.8904879]
2. Fan H, Smith SP, Lederman R, Chang S. Why people trust in online health communities: an integrated approach. In: Proceedings of the 21st Australasian Conference on Information Systems. 2010 Presented at: ACIS '10; August 18-20, 2010; Yamagata, Japan.
3. Kim HS, Mrotek A. A functional and structural diagnosis of online health communities sustainability: a focus on resource richness and site design features. Comput Human Behav 2016 Oct;63:362-372. [doi: 10.1016/j.chb.2016.05.004]
4. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. Patient Educ Couns 2006 Aug;62(2):178-188. [doi: 10.1016/j.pec.2005.10.003] [Medline: 16406472]

5. Nasralah T, Noteboom C, Wahbeh A, Al-Ramahi MA. Online health recommendation system: a social support perspective. Dakota State University. 2017. URL: https://scholar.dsu.edu/cgi/viewcontent.cgi?article=1008&context=bispapers [accessed 2022-06-10]
6. Ji X, Geller J, Chun S. Social InfoButtons for patient-oriented healthcare knowledge support. In: Proceedings of the 2014 AMIA Annual Symposium. 2014 Presented at: AMIA '14; November 15-19, 2014; Washington, DC, USA.
7. Waterworth S, Luker KA. Reluctant collaborators: do patients want to be involved in decisions concerning care? J Adv Nurs 1990 Aug;15(8):971-976. [doi: 10.1111/j.1365-2648.1990.tb01953.x] [Medline: 2229694]
8. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
9. Breast Cancer. Cancer Survivors Network. URL: https://csn.cancer.org/forum/127 [accessed 2022-06-10]
10. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: controlled experimental survey study. J Med Internet Res 2018 Mar 26;20(3):e99 [FREE Full text] [doi: 10.2196/jmir.8986] [Medline: 29581091]
11. Zhang D, Yue WT. Social media use in decision making: special issue of decision support systems for the 10th workshop on e-business. Decis Support Syst 2014 Jul;63:65-66. [doi: 10.1016/j.dss.2013.08.007]
12. Wu B, Jiang S, Chen H. Effects of individuals' motivations on communications in online health forums. Soc Behav Pers 2016 Mar 23;44(2):299-312. [doi: 10.2224/sbp.2016.44.2.299]
13. Pan W, Shen C, Feng B. You get what you give: understanding reply reciprocity and social capital in online health support forums. J Health Commun 2017 Jan;22(1):45-52. [doi: 10.1080/10810730.2016.1250845] [Medline: 28027009]
14. Li S, Feng B, Wingate VS. Give thanks for a little and you will find a lot: the role of a support seeker's reply in online support provision. Commun Monogr 2019;86(2):251-270. [doi: 10.1080/03637751.2018.1539237]
15. Yu J, Hu Y, Yu M, Di Z. Analyzing netizens' view and reply behaviors on the forum. Phys A Stat Mech Appl 2010 Aug;389(16):3267-3273. [doi: 10.1016/j.physa.2010.03.043]
16. Tang X, Yang CC. Identifying influential users in an online healthcare social network. In: Proceedings of the 2010 IEEE International Conference on Intelligence and Security Informatics. 2010 Presented at: ISI '10; May 23-26, 2010; Vancouver, Canada p. 43-48. [doi: 10.1109/ISI.2010.5484779]
17. Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in Twitter: the million follower fallacy. Proc Conf AAAI Artif Intell 2010;4(1):10-17.
18. Ishfaq U, Khan HU, Iqbal K. Identifying the influential bloggers: a modular approach based on sentiment analysis. J Web Eng 2017;16(5-6):505-523.
19. Zhao K, Qiu B, Caragea C, Wu D, Mitra P, Yen J, et al. Identifying leaders in an online cancer survivor community. In: Proceedings of the 21st Annual Workshop on Information Technologies and Systems. 2011 Presented at: WITS '11; December 3-4, 2011; Shanghai, China p. 115-120.
20. Hou X, Lei CU, Kwok YK. OP-DCI: a riskless K-means clustering for influential user identification in MOOC forum. In: Proceedings of the 16th IEEE International Conference on Machine Learning and Applications. 2017 Presented at: ICMLA '17; December 18-21, 2017; Cancun, Mexico p. 936-939. [doi: 10.1109/icmla.2017.00-34]
21. Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. J Am Med Inform Assoc 2014 Oct;21(e2):e212-e218 [FREE Full text] [doi: 10.1136/amiajnl-2013-002282] [Medline: 24449805]
22. Cercel DC, Trausan-Matu S. Opinion influence analysis in online forum threads. In: Proceedings of the 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. 2014 Presented at: SYNASC '14; September 22-25, 2014; Timisoara, Romania p. 228-235. [doi: 10.1109/synasc.2014.38]
23. Qiu B, Zhao K, Mitra P, Wu D, Caragea C, Yen J, et al. Get online support, feel better--sentiment analysis and dynamics in an online cancer survivor community. In: Proceedings of the IEEE 3rd International Conference on Privacy, Security, Risk and Trust and IEEE 3rd International Conference on Social Computing. 2011 Presented at: SocialCom '11; October 9-11, 2011; Boston, MA, USA p. 274-281. [doi: 10.1109/PASSAT/SocialCom.2011.127]
24. Definition of influence. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/influence [accessed 2022-06-10]
25. Kidd C, Hayden BY. The psychology and neuroscience of curiosity. Neuron 2015 Nov 04;88(3):449-460 [FREE Full text] [doi: 10.1016/j.neuron.2015.09.010] [Medline: 26539887]
26. Palloff RM, Pratt K. Lessons from the Virtual Classroom: The Realities of Online Teaching. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; 2013.
27. Adjei MT, Noble SM, Noble CH. The influence of C2C communications in online brand communities on customer purchase behavior. J Acad Mark Sci 2009 Nov 13;38(5):634-653. [doi: 10.1007/s11747-009-0178-5]
28. Liu Y, Chen F, Chen Y. Learning thread reply structure on patient forums. In: Proceedings of the 2013 International Workshop on Data Management & Analytics for Healthcare. 2013 Presented at: DARE '13; November 1, 2013; San Francisco, CA, USA p. 1-4. [doi: 10.1145/2512410.2512426]
29. Huang PS, He X, Gao J, Deng L, Acero A, Heck L. Learning deep structured semantic models for web search using clickthrough data. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management.

2013 Presented at: CIKM '13; October 27-November 1, 2013; San Francisco, CA, USA p. 2333-2338. [doi: 10.1145/2505515.2505665]
30. Shen Y, He X, Gao J, Deng L, Mesnil G. Learning semantic representations using convolutional neural networks for Web search. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW '14 Companion; April 7-11, 2014; Seoul, South Korea p. 373-374. [doi: 10.1145/2567948.2577348]
31. Hu B, Lu Z, Li H, Chen Q. Convolutional neural network architectures for matching natural language sentences. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. 2014 Presented at: NIPS '14; December 8-13, 2014; Montreal, Canada p. 2042-2050.
32. Wang YC, Kraut R, Levine JM. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Seattle, WA, USA p. 833-842. [doi: 10.1145/2145204.2145329]
33. Lu Z, Li H. A deep architecture for matching short texts. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. 2013 Presented at: NIPS '13; December 5-10, 2013; Lake Tahoe, NV, USA p. 1367-1375. [doi: 10.7551/mitpress/11474.003.0014]
34. Pang L, Lan Y, Guo J, Xu J, Wan S, Cheng X. Text matching as image recognition. Proc Conf AAAI Artif Intell 2016 Mar 5;30(1):10341. [doi: 10.1609/aaai.v30i1.10341]
35. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv 2018 Oct 11. [doi: 10.48550/arXiv.1810.04805]
36. Efron M, Winget M. Questions are content: a taxonomy of questions in a microblogging environment. Proc Am Soc Info Sci Tech 2011 Feb 03;47(1):1-10. [doi: 10.1002/meet.14504701208]
37. Liu Z, Jansen BJ. Questioner or question: predicting the response rate in social question and answering on Sina Weibo. Inf Process Manag 2018 Mar;54(2):159-174. [doi: 10.1016/j.ipm.2017.10.004]
38. Ranganath S, Hu X, Tang J, Wang S, Liu H. Identifying rhetorical questions in social media. Proc Conf AAAI Artif Intell 2016;10(1):667-670.
39. Ranganath S, Hu X, Tang J, Wang S, Liu H. Understanding and identifying rhetorical questions in social media. ACM Trans Intell Syst Technol 2018 Mar 31;9(2):1-22. [doi: 10.1145/3108364]
40. Ojokoh B, Igbe T, Araoye A, Ameh F. Question identification and classification on an academic question answering site. In: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries. 2016 Presented at: JCDL '16; June 19-23, 2016; Newark, NJ, USA p. 223-224. [doi: 10.1145/2910896.2925442]
41. Bird S, Loper E. NLTK: the natural language toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004 Presented at: ACL '04; July 21-26, 2004; Barcelona, Spain p. 214-217.
42. Goodfellow I, Yoshua B, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; Nov 2016:4.
43. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. 2015 Presented at: ICLR '15; May 7-9, 2015; San Diego, CA, USA.
44. Mingda Li, Jinhe Shi, Yi Chen. NJIT-AI-in-Healthcare / Influence-Relationship-Classification. GitHub. 2022 Feb 27. URL: https://github.com/NJIT-AI-in-Healthcare/Influence-Relationship-Classification [accessed 2022-06-10]
45. Scrapy. URL: https://scrapy.org/ [accessed 2022-06-10]
46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 Jun;60(6):84-90. [doi: 10.1145/3065386]
47. Liu Y, Shi J, Chen Y. Patient-centered and experience-aware mining for effective adverse drug reaction discovery in online health forums. J Assoc Inf Sci Technol 2017 Sep 22;69(2):215-228. [doi: 10.1002/asi.23929]

Abbreviations

ARC-I: Architecture-I
ARC-II: Architecture-II
BERT: Bidirectional Encoder Representations from Transformers
CSN: Cancer Survivors Network
DSSM: Deep Structured Semantic Models
OHC: online health community
PR AUC: area under the precision-recall curve
ROC AUC: area under the receiver operating characteristic curve

Edited by R Kukafka; submitted 03.06.21; peer-reviewed by S Liu, G Lim; comments to author 02.08.21; revised version received 05.10.21; accepted 21.03.22; published 31.08.22

Please cite as:
Li M, Shi J, Chen Y
Identifying Influences in Patient Decision-making Processes in Online Health Communities: Data Science Approach
J Med Internet Res 2022;24(8):e30634
URL: https://www.jmir.org/2022/8/e30634
doi: 10.2196/30634
PMID:

Original Paper

The Impact of a Place-Tailored Digital Health App Promoting Exercise Classes on African American Women's Physical Activity and Obesity: Simulation Study

Tiffany M Powell-Wiley[1,2], MD, MPH; Marie F Martinez[3,4], MSPH; Kosuke Tamura[5], PhD; Sam J Neally[1], BA; Kelly J O'Shea[3,4], BSFS; Kaveri Curlin[1], BS; Yardley Albarracin[3], MPH; Nithya P Vijayakumar[1], MD; Matthew Morgan[3], MBE; Erika Ortiz-Chaparro[1], MD; Sarah M Bartsch[3,4], MPH; Foster Osei Baah[1], PhD; Patrick T Wedlock[3,4], MSPH; Lola R Ortiz-Whittingham[1], BS; Sheryl Scannell[3,4], MS; Kameswari A Potharaju[1], BA; Samuel Randall[3], BS; Mario Solano Gonzales[3], BS; Molly Domino[3], BS; Kushi Ranganath[3], BS; Daniel Hertenstein[3], BS; Rafay Syed[3], MS; Colleen Weatherwax[3,4], MS; Bruce Y Lee[3,4], MD, MBA

[1]Social Determinants of Obesity and Cardiovascular Risk Laboratory, Cardiovascular Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, United States

[2]Intramural Research Program, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

[3]Public Health Informatics, Computational, and Operations Research, CUNY Graduate School of Public Health and Health Policy, New York, NY, United States

[4]Center for Advanced Technology and Communication in Health, CUNY Graduate School of Public Health and Health Policy, New York, NY, United States

[5]Socio-Spatial Determinants of Health (SSDH) Laboratory, Population and Community Sciences Branch, Intramural Research Program, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

Corresponding Author:
Tiffany M Powell-Wiley, MD, MPH
Social Determinants of Obesity and Cardiovascular Risk Laboratory
Cardiovascular Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute
National Institutes of Health
10 Center Drive
Bethesda, MD, 20892
United States
Phone: 1 (301) 547 5275
Email: tiffany.powell-wiley@nih.gov

Abstract

Background: The increasing prevalence of smartphone apps to help people find different services raises the question of whether apps to help people find physical activity (PA) locations would help better prevent and control having overweight or obesity.

Objective: The aim of this paper is to determine and quantify the potential impact of a digital health intervention for African American women prior to allocating financial resources toward implementation.

Methods: We developed our Virtual Population Obesity Prevention, agent-based model of Washington, DC, to simulate the impact of a place-tailored digital health app that provides information about free recreation center classes on PA, BMI, and overweight and obesity prevalence among African American women.

Results: When the app is introduced at the beginning of the simulation, with app engagement at 25% (eg, 25% [41,839/167,356] of women aware of the app; 25% [10,460/41,839] of those aware downloading the app; and 25% [2615/10,460] of those who download it receiving regular push notifications), and a 25% (25/100) baseline probability to exercise (eg, without the app), there are no statistically significant increases in PA levels or decreases in BMI or obesity prevalence over 5 years across the population. When 50% (83,678/167,356) of women are aware of the app; 58.23% (48,725/83,678) of those who are aware download it; and 55% (26,799/48,725) of those who download it receive regular push notifications, in line with existing studies on app usage, introducing the app on average increases PA and decreases weight or obesity prevalence, though the changes are not statistically significant. When app engagement increased to 75% (125,517/167,356) of women who were aware, 75% (94,138/125,517) of those who were aware downloading it, and 75% (70,603/94,138) of those who downloaded it opting into the app's push notifications, there were statistically significant changes in PA participation, minutes of PA and obesity prevalence.

Conclusions: Our study shows that a digital health app that helps identify recreation center classes does not result in substantive population-wide health effects at lower levels of app engagement. For the app to result in statistically significant increases in PA and reductions in obesity prevalence over 5 years, there needs to be at least 75% (125,517/167,356) of women aware of the app, 75% (94,138/125,517) of those aware of the app download it, and 75% (70,603/94,138) of those who download it opt into push notifications. Nevertheless, the app cannot fully overcome lack of access to recreation centers; therefore, public health administrators as well as parks and recreation agencies might consider incorporating this type of technology into multilevel interventions that also target the built environment and other social determinants of health.

(J Med Internet Res 2022;24(8):e30581) doi: 10.2196/30581

KEYWORDS

computational modeling; digital health; physical activity; BMI; obesity; built environment; impact; app; exercise; simulation; intervention; women; African American; agent

Introduction

The increasing prevalence of smartphone apps to help people find different services (eg, Yelp and OpenTable to find restaurants, Fandango to find movie theaters, AllTrails to find hikes, GasBuddy to find gas stations, Expedia to find hotels, and Zillow to find homes and apartments) raises the question of whether apps to help people find physical activity (PA) locations (eg, ClassPass [1] and Fit Reserve [2]) would help to better prevent and control having overweight and obesity. Such place-tailored apps can help assemble, collate, and present information that may be available on different websites so that an individual can quickly find the closest location of interest. These place-tailored apps can be particularly helpful for PA locations and opportunities since they may exist in different and less obvious forms (eg, irregular timing of classes, walking and bike paths, outdoor tracks, and tennis or basketball courts). Such an app can also offer crowdsourced ratings of each location, details about specific services (eg, time, availability, costs, promotions, and deals), and even social connections with people who have the same interests or are in the same area. Previous studies have shown that people may not be aware of or have difficulty finding locations to engage in PA [3-5]. This may be the case in underresourced and otherwise disadvantaged communities where parks, affordable gyms, and other opportunities may be more difficult to find if they are in less-frequented or obscure locations, or if they are not regularly advertised or promoted [6]. African American women who live disproportionately in underresourced communities spend at least as much time as any other racial or ethnic group using apps and the internet (approximately 19 hours and 27 minutes each week versus 17 hours and 8 minutes each week), and approximately 80% of African American women own a smartphone [7], raising the possibility that this could be an effective means to help these women find PA opportunities. However, before such an app is rolled out in the "real world," it can be helpful to use simulation modeling to guide the design and test the potential impact of such an app. Such an approach is used in other fields (eg, aeronautical engineering and manufacturing) since running simulation models can take much less time and can be significantly less costly than conducting a real-world trial (which can take months to set up, recruit for, and implement). Moreover, once a trial is completed, one cannot go back and change the circumstances as they can in a simulation model. Therefore, we further developed our agent-based simulation model of Washington, DC to test the impact of such a place-tailored digital health app.

Methods

Ethics Approval

All authors' institutions were included in the institutional review board approval (IRB #00004203) at Johns Hopkins as the study began while certain members of the research team (MCF, KJO, YA, MM, SMB, PTW, SS, SR, MSG, MD, KR, DH, RS, and BYL) were based at Johns Hopkins.

Model of Washington, DC

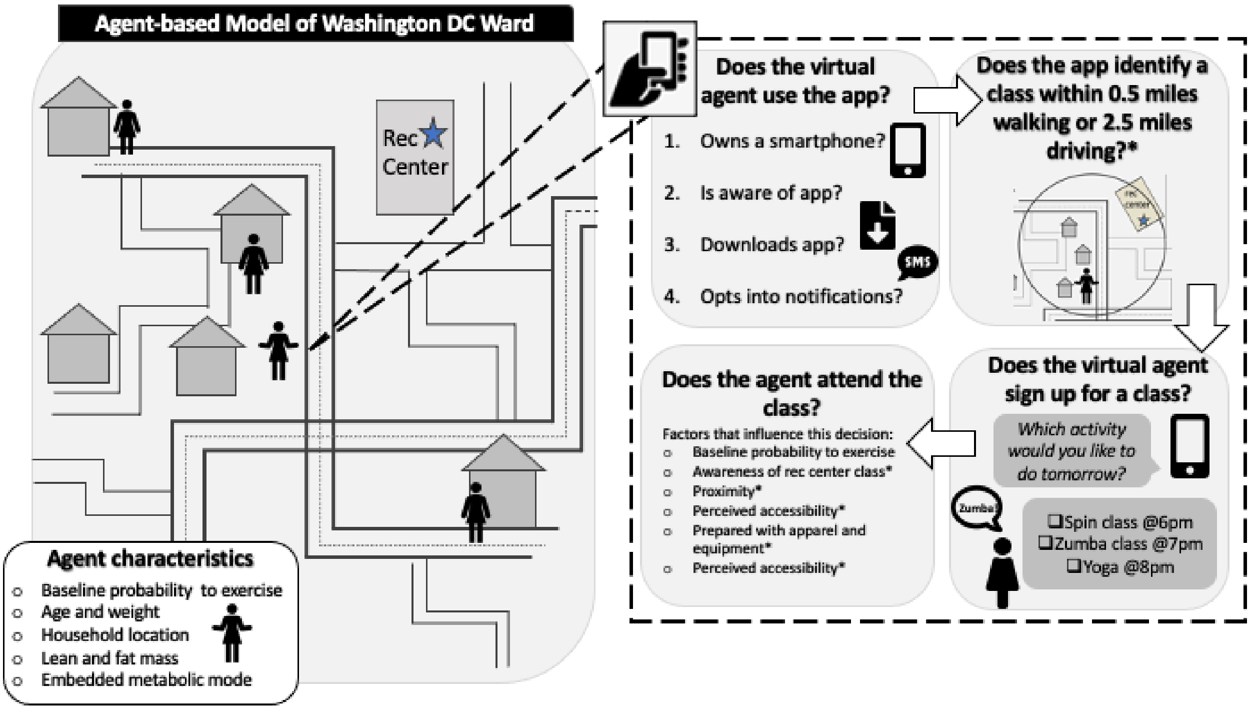
We used and further developed a Virtual Population Obesity Prevention, agent-based model of Washington, DC in 2020-2021 [8,9], which includes computer model–based representations of households, workplaces, and recreation centers throughout all 8 wards (similar to districts in other cities) in Washington, DC.

Agents Representing People

We represented each of the 167,356 African American women (aged 18-65 years) living in Washington, DC with a computer model–based agent. Each agent (ie, each African American woman in Washington, DC) has attributes for age, height, lean or fat mass, household location, work location, and income based on representative data for the region and population. Each agent also has an embedded metabolic model, which converts daily caloric intake and expenditure to corresponding lean or fat mass [10,11]. Caloric expenditure from exercise is determined by exercise intensity, duration, and the agent's current body weight [10,11]. Since individuals may vary in their inclination to exercise, each agent had a baseline probability of wanting to exercise each day. This accounts for an agent's past experiences and existing tendencies to exercise and includes factors such as household financial status, family responsibilities, chronic health conditions, and social influences. Different scenarios ranged this baseline probability from 10% (10/100) to 50% (50/100) to explore how this probability might affect the results.

In each simulated day, women may participate in a recreation center class, depending on a number of factors (Figure 1; Multimedia Appendix 1, Table S1 [8, 9, 12-23]), including the following: (1) her baseline probability to exercise (this accounts for an agent's past experience and existing tendencies to participate in recreation center classes), which we vary between simulation experiments; (2) objective accessibility to locations, based on the geographic locations of recreation centers [12], the distance individuals need to travel to reach these locations, and access to the types of transportation (eg, cars) that might be required to reach locations further away [24,25]; (3) perceived accessibility of locations [15], based on the individual's understanding and knowledge of nearby recreation centers; (4) awareness of classes at recreation centers; and (5) preparedness to exercise (whether or not she remembers her apparel and equipment).

Figure 1. A digital health app that helps locate and send reminders about recreation (rec) center classes. *Factors influenced by phone app.



Representations of Recreation Centers

Multimedia Appendix 1, Table S2 shows key characteristics (eg, number of recreation centers) for each ward. If an agent ultimately participates in a recreation center class, she is active for 50 minutes [12] at an intensity of approximately 6.5 metabolic equivalents [16].

Representations of Digital Health App

In the model, we represent a digital health app that helps locate and send reminders about in-person recreation center classes to increase the agents' likelihood of participation (Figure 1). This mobile app uses a database of public locations that have been previously identified as locations for PA, such as recreation centers in the case of this paper. Once this registry is verified, a geofence—a geographic boundary—can be created within the mobile app with a set distance surrounding the chosen location; in this case, our simulated app searched for recreation centers within 0.5 miles of the user. When the simulated mobile app detects that the user is within this defined boundary, it will generate a notification that will alert the user of the available resources in the area. Unlike existing fitness apps, this digital health app considers the geographic location of the user and the recreation centers to connect agents with recreation center classes that align with their neighborhood environment and schedules; prompts users to remind them about upcoming classes and what equipment they will need; and provides individually tailored information about class time, location, and necessary equipment to maximize user engagement (Figure 1). If an agent has a smartphone, downloads the app, and opts into notifications (Multimedia Appendix 1, Table S1), the app will send a question each evening asking the user which activity or class she would like to participate in the following day, thereby increasing an agent's knowledge of class schedules. After selecting the class, agents will receive a notification with a reminder of the class's time, location, and activity, as well as a reminder to bring clothes or equipment, thereby increasing an agent's probability of being prepared for and attending class. When representing the digital health app, we introduced it at the beginning of the 5-year simulation, but not all participants continued to use the app for the entire simulation duration (eg, we represented attrition, people discontinuing app use, during the 3 months following the introduction of the app; Multimedia Appendix 1, Table S1).

Representations of Engagement With the Digital Health App

Since only a certain percentage of the population may be aware that the app is available, we varied the proportion of women across the population who, in a given scenario, were aware of the app, subsequently downloaded the app, and then opted into push notifications (25%-75%). This means, 25% (41,839/167,356) of women are aware of the app, 25%

(10,460/41,841) of those who are aware download it, and 25% (2615/10,460) of those who download it receive regular push notifications from the app. We ranged this to 75% (125,517/167,356) of women aware of the app, 75% (94,138/125,517) of those who are aware download it, and 75% (70,603/94,138) of those who download it opt into the app's push notifications. Varying the level of user engagement across a range can help identify the thresholds of app engagement that result in observable and statistically significant impacts on PA and weight.

Simulation Experiments

We used the model of Washington, DC to simulate the impact of a digital health app on in-person recreation center class participation, recreation center class PA (minutes per week), subsequent changes in BMI, as well as the prevalence of obesity and the state of having overweight. Each simulation experiment consisted of running the model of Washington, DC and all 167,356 computer model–based agents, 10 times over 5 simulated years.

Validation

Validation consisted of comparing different model-generated metrics to observed values to determine if the model was representing what was occurring. For example, when we ran simulation runs, we saw that, on average, 2.1% (3514/167,356) of women were participating in recreation center classes daily compared to the observed 3.8% from the 2017 American Time Use Survey [17]. Since the people who exercised on one day will not necessarily be the same people who exercised on a different day, there will be a certain proportion of the population that exercised at least once over the course of the month. Thus, we also simulated the average percentage of women participating in recreation center classes at least once on a monthly basis (19.1% [31,965/176,356]) and compared this to the observed proportion of women participating in workout class activity on a monthly basis (16.1%), as reported by the Behavioral Risk Factors Surveillance System [18]. The model-generated data generally matched the observed data, and the differences are likely due to differences between populations and the classes available to that population. Further model validation details are available in Multimedia Appendix 1.

Results

No Mobile App

Table 1 shows PA from recreation center classes and weight-related outcomes after 5 years with no mobile app for different baseline probabilities to exercise. Figure 2 shows how the percent of women who exercised at least once in the simulation when there was no app varied by the Washington, DC ward. For example, Ward 6 had the highest percent of the population who exercised at least once (69.1% [4331/20,739], 95% CI 68.9%-69.2%), while Ward 7 had the lowest (48% [15,710/32,729], 95% CI 47.9%-48.0%) when there was no mobile app (25% [25/100] baseline probability to exercise). This trend in ward-level variation was consistent across all baseline exercise probabilities.

Table 1. Physical activity, overweight, obesity, BMI outcomes by baseline probability to exercise for different scenarios (eg, with and without digital health app).

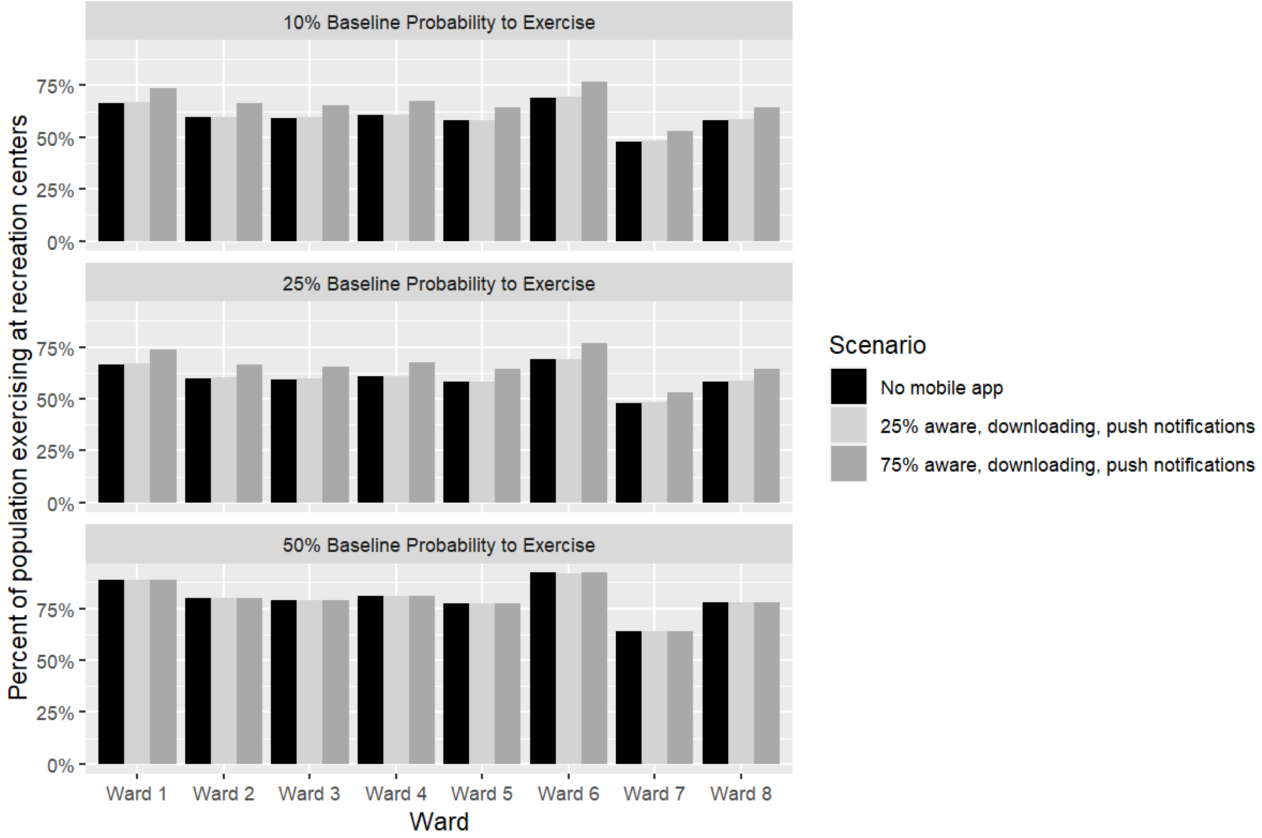
Simulation scenarios at each baseline probability to exercise	Percent of population exercising at recreation centers, mean (95% CI)	Average number of physical activity min/week, mean (95% CI)	Overweight prevalence, mean (95% CI)	Obesity prevalence, mean (95% CI)	Average BMI, mean (95% CI)	Average BMI among women with obesity, mean (95% CI)
10% (10/100) baseline probability to exercise						
No digital health app	58.66 (54.65-62.67)	36.97 (34.45-39.50)	24.44 (23.91-24.97)	56.10 (54.56-57.64)	30.16 (29.86-30.45)	34.20 (34.00-34.41)
Introducing place-tailored digital health app						
25%-25%-25%[a]	58.91 (54.87-62.94)	37.26 (34.71-39.81)	24.42 (23.88-24.96)	56.09 (54.53-57.65)	30.15 (29.86-30.45)	34.21 (34.00-34.43)
50%-50%-50%[b]	61.09 (56.92-65.26)	39.83 (37.12-42.54)	24.45 (23.91-24.98)	55.67 (54.15-57.19)	30.07 (29.78-30.36)	34.16 (33.94-34.37)
75%-75%-75%[c]	65.10 (60.64-69.56)	44.45 (41.41-47.50)	24.70 (24.21-25.20)	54.68 (53.12-56.25)	29.90 (29.60-30.19)	34.04 (33.83-34.26)
25% (25/100) baseline probability to exercise						
No digital health app	58.67 (54.66-62.68)	52.84 (49.23-56.45)	25.52 (25.04-26.01)	52.75 (51.06-54.43)	29.56 (29.27-29.86)	33.81 (33.61-34.01)
Introducing place-tailored digital health app						
25%-25%-25%[a]	58.92 (54.89-62.94)	53.25 (49.61-56.89)	25.54 (25.04-26.04)	52.62 (50.91-54.33)	29.56 (29.26-29.86)	33.83 (33.62-34.05)
50%-50%-50%[b]	61.17 (56.99-65.35)	56.98 (53.09-60.88)	26.24 (25.68-26.80)	51.25 (49.47-53.04)	29.44 (29.14-29.74)	33.83 (33.62-34.03)
75%-75%-75%[c]	65.10 (60.64-69.55)	63.52 (59.18-67.87)	27.72 (27.05-28.40)	48.66 (46.75-50.56)	29.23 (28.92-29.54)	33.82 (33.63-34.02)
50% (50/100) baseline probability to exercise						
No digital health app	78.30 (72.96-83.64)	86.33 (80.43-92.22)	27.88 (26.36-29.39)	44.42 (41.63-47.20)	28.38 (28.06-28.70)	33.00 (32.80-33.21)
Introducing place-tailored digital health app						
25%-25%-25%[a]	78.22 (72.86-83.58)	86.88 (80.93-92.83)	28.24 (26.81-29.67)	43.90 (41.27-46.52)	28.38 (28.05-28.71)	33.08 (32.84-33.31)
50%-50%-50%[b]	78.25 (72.90-83.60)	92.17 (85.89-98.45)	28.57 (27.17-29.96)	42.63 (40.03-45.23)	28.24 (27.91-28.57)	33.10 (32.86-33.34)
75%-75%-75%[c]	78.29 (72.95-83.63)	101.41 (94.48-108.33)	29.40 (28.15-30.66)	40.27 (37.75-42.78)	28.00 (27.65-28.34)	33.15 (32.91-33.40)

[a] 25% (41,839/167,356) aware of the app, 25% (10,460/41,839) of those who are aware download the app, and 25% (2615/10,460) of those who download it receive notifications.

[b] 50% (83,678/167,356) aware of the app, 50% (48,725/83,678) of those who are aware download the app, and 50% (26,799/48,725) of those who download it receive notifications.

[c] 75% (125,517/167,356) aware of the app, 75% (94,138/125,517) of those who are aware download app, and 75% (70,603/94,138) of those who download it receive app notifications.

Figure 2. Percent of women exercising with and without the mobile app within each ward in Washington, DC.



Impact of Introducing a Place-Tailored Mobile App That Connects Users to Recreation Center Classes

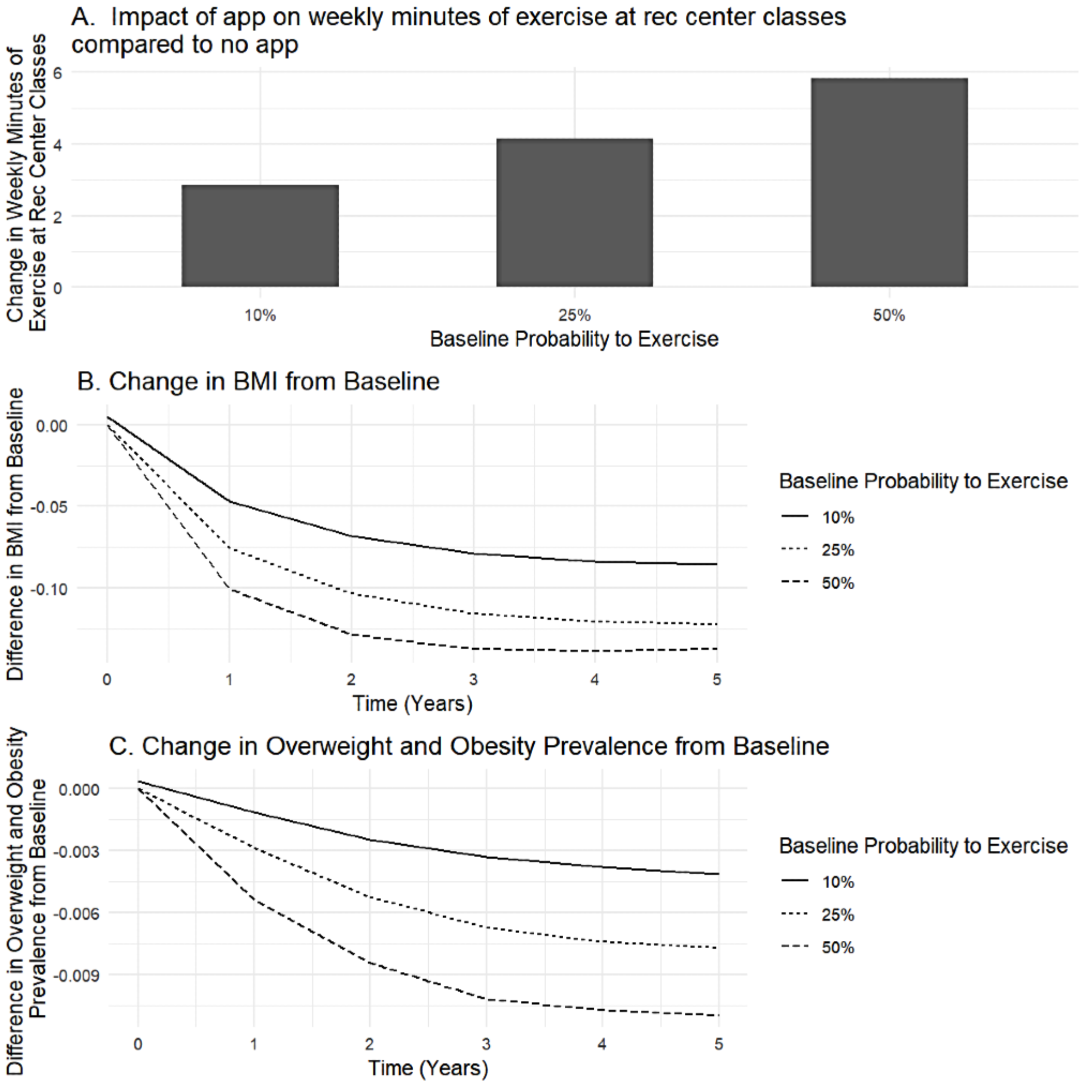
With lower levels of user engagement with the mobile phone app, that is 25% aware of app (41,839/167,356), 25% of those aware download app (10,460/41,839), and 25% (2615/10,460) of those who download it receive app notifications, the app had a negligible and nonsignificant impact on the additional minutes of PA (<1 minute), on the additional percent of women who ever exercise (0.2% [335/167,356]; Figure 2), and on reductions in obesity prevalence (0.1% [167/167,356]). Thus, even lower levels of app engagement (eg, below 25% [25/100]) would have no effect on physical activity and weight.

Increasing user engagement to approximately 50% (eg, 50% aware [83,678/167,356], 58.23% [48,725/83,678] of those who are aware download the app [19], and 55% [26,799/48,725] of those who download it receive regular push notifications [20]) resulted in moderate improvements to PA from recreation center classes and weight-related outcomes across the population. Figure 3 shows these observable changes to PA (panel A), BMI (panel B), and overweight and obesity prevalence (panel C). With a 10% (10/100) baseline probability of exercise, the PA minutes per week increase by 2.9 minutes (95% CI –1.4 to 17.9), BMI decreases by 0.09 kg/m^2 (95% CI –0.56 to 0.39), and obesity prevalence decreases by an absolute 0.43% (720/167,356; 95% CI –2.7% to 2.93%) at the end of the 5-year simulation. When baseline probability increases to 50% (50/100), there are larger increases in weekly PA minutes (5.4 minutes, 95% CI –4.1 to 15.8), and larger reductions in BMI (0.14 kg/m^2, 95% CI –0.67 to –0.4) and obesity prevalence (1.8% [3012/167,356]; 95% CI –2.6% to 6.2%).

The percent of women attending at least one recreation center class over the course of the simulation shows additional gains between when the baseline probability to exercise is between 10% (10/100; 2.43% [4067/167,356], 95% CI –4.24% to 9.1%) and 25% (25/100; 2.5% [4184/167,356]; 95% CI –4.2% to 9.2%). When the baseline probability to exercise is 50% (50/100), the percent of women exercising at least once hits a ceiling of 78% (130,538/167,356) (increase of 0.05% [84/167,356]; 95% CI –8.68% to 8.77%), due to the location and accessibility of recreation centers for some women. Thus, at lower probabilities to exercise (eg, 10%-25%), the app is more effective at increasing the number of women participating in at least one recreation center class (Figure 2). However, additional PA minutes per week from recreation center classes increase with baseline probability to exercise (eg, 4.14, 95% CI –1.9 to 10.2 vs 5.9, 95% CI –4.1 to 15.7 minutes per week at 25% [25/100] and 50% [50/100] baseline probabilities to exercise, respectively; Figure 3). Figure 3 also shows how reductions in BMI and overweight and obesity prevalence due to app use accrue over time during the 5-year simulation.

Further increasing app engagement to 75%, with 75% (125,517/167,356) of women aware of the app, 75% (94,138/125,517) of those who are aware downloading the app, and 75% (70,603/94,138) of those who download it opting into the app's push notifications resulted in statistically significant gains to PA and reductions in obesity prevalence. For example, weekly PA increased by 10.7 (95% CI 4.2-17.2) minutes per week, and obesity prevalence decreased by an absolute 4.09% (6,845/167,356; 95% CI 1.2%-7.0%) with 25% baseline exercise probability (Table 1).

Figure 3. Impact of mobile app on physical activity, BMI, as well as overweight and obesity prevalence at each baseline probability to exercise. Rec: recreation.



Ward-Level Impact of Place-Tailored Mobile App

The results varied substantially by ward. For example, at 25% (25/100) baseline probability to exercise (assuming 50% [83,678/167,356] aware, 50% [48,725/83,678] of those who are aware downloading the app, and 50% [26,799/48,725] of those who download it receiving app notifications), Ward 6 had the highest absolute increase in average PA minutes per week (4.85, 95% CI 4.58-5.11), and the greatest reduction in average BMI (–0.15 kg/m^2, 95% CI –0.19 to –0.11). However, Ward 7 had the lowest (3.39, 95% CI 3.24-3.53) increase in PA minutes per week and the smallest reduction in BMI (–0.09 kg/m^2; 95% CI –0.12 to –0.06). Changes in overweight and obesity prevalence also varied between wards and decreased by as much as 2.6% (539/20,739; 95% CI 2.3%-2.9%) in Ward 6, where participation in recreation center classes was highest and as little as 1.9% (622/32,729; 95% CI 1.7%-2.1%) in Ward 7 (25% baseline exercise probability).

Discussion

Principal Findings

Our simulation model of African American women in Washington, DC, and their use of a place-tailored digital health app to help identify recreation center classes shows that the app does not result in substantive population-wide health effects at lower levels of app engagement (eg, 25% of women are aware of the app, 25% of those aware of the app download it, and 25% of those who download it receive regular push notifications from the app). When 50% of women are aware of the app, 58.23% of those who are aware download the app, and 55% of those who download it receive regular push notifications from the app, there are observable changes in PA and weight across the population, but the impact is not statistically significant. For the app to result in statistically significant increases in PA and reductions to obesity prevalence over 5 years, there needs to be at least 75% of women who are aware of the app, 75% of those aware of the app downloading it, and 75% of those who download it opting into the app's push notifications. Thus, we demonstrated the minimum levels of engagement needed at the outset of a mobile phone app campaign (approximately 50% aware of the app, 50% of those who are aware download the app, and 50% of those who download it receive app notifications, assuming reductions in use over the first 3 months) to observe a change in PA and weight across the population. Studies have shown how perceived usefulness of an app, user-friendliness, backing from health care professionals, and continued engagement impact app usage [26,27] could be addressed through a structured marketing and communications strategy. Thus, future interventions should prioritize efforts to increase marketing for the place-tailored app to increase the percent of women who are aware of and use the app to reach the impactful threshold of engagement and obtain further benefits.

Further, our results show that a place-tailored app is more likely to be successful in increasing PA in those who already have a higher likelihood to exercise. While the results showed that the app was successful at encouraging individuals who have a low baseline probability (eg, 10% [10/100] and 25% [25/100]) to exercise to attend at least one new class over the course of the simulated period, this alone was not enough to drive a sustained change in regular exercise. The app did a better job at increasing the average duration of PA each week as baseline probability to exercise increased. This indicates that improving knowledge of recreation center classes, while important, should be coupled with interventions to help overcome personal and social barriers (eg, limited social support for PA or time constraints) that determine baseline exercise probability [28,29]. Place-tailored digital health apps could potentially address some of these barriers through the release of new features and functionality such as a social networking component [30,31].

Regardless of user engagement with the app, place-tailored digital health apps need to be combined with increasing physical access to recreation centers to see greater than additive effects in PA and subsequent health outcomes. There is a limit to a place-tailored app's impact because some individuals cannot access recreation centers due to the distance and lack of transportation (eg, access to car) from their home location. As shown in our results, there are clear disparities in the success of the app in improving health outcomes in neighborhoods with greater access to recreation centers (with nearly a 1.4-fold increase in the use of recreation center classes in these neighborhoods [eg, Ward 6]) compared to neighborhoods with less accessible recreation centers (eg, Ward 7), even with 75% of women who are aware of the app, 75% of those aware of the app download it, and 75% of those who download it opt into the app's push notifications. Past studies have shown that lower-income neighborhoods in many cities around the United States have less accessible PA locations and recreation centers [32]. Therefore, it is important for public health administrators and park and recreation agencies to consider pairing this type of digital health technology with improvements to recreation center access such as changes to the built environment, perceived safety, or transportation.

Our results also show that it takes time for the effect of the place-tailored mobile app to fully manifest (>2 years). In general, 1 year is not enough time to see an impact on BMI and overweight and obesity prevalence, as population-level effects on weight and subsequent health benefits accrue over years. This shows the need to continuously measure the value of intervention programs over a period of several years, since reductions in overweight and obesity prevalence may not be demonstrated immediately, and effective interventions may wrongly be deemed unsuccessful if evaluated too early. Accounting for this ramp-up period is important, as it can also take time for a new technology to be adopted and used. Our results show that the speed of the reduction in overweight and obesity prevalence in the population increases year after year as adoption rates increase, revealing a potential opportunity to increase momentum as more users adopt similar place-tailored digital health technology.

In addition to being able to simulate extended periods of time, another benefit of simulation modeling is that it can be adapted and refined over time. For example, simulation modeling can be used in conjunction with clinical trials [33,34] so that the model can continuously inform digital health phone app design and multipronged PA interventions. The simulation model can be run first, to help determine the impact of an app, which can then inform the implementation of a trial. Data and information from the trial can then further update the model. This iterative process can continue until the app or intervention is optimally designed.

Limitations

All models are simplifications of reality and cannot account for all possible factors that may affect PA decision-making. Our model included a few simplifying assumptions. For example, we did not account for objective accessibility to a recreation center near a woman's workplace and used the objective accessibility near the home as a proxy. In addition, since we wanted to demonstrate how to design an app that harnesses geographic location and the value of such an app, our study focused on the app locating and reminding individuals about in-person classes, rather than web-based classes. However, such an app may offer similar benefits for web-based classes such as reminding individuals about when classes are scheduled and what equipment is needed, while reducing potential geographic barriers to exercise. We also assumed that in-person classes are available (eg, not during a public health emergency such as the COVID-19 pandemic). When determining body weight changes for each woman, we assumed that compensatory eating did not occur. Our model simulated behavior of and used data specific to Washington, DC African American women, which may limit generalizability to other populations or geographic areas.

Conclusions

Our study shows that a digital health app that helps identify recreation center classes does not result in substantive population-wide health effects at lower levels of app engagement (eg, 25% of women who are aware of the app, 25% of those who are aware of the app download it, and 25% of those who download it receive regular push notifications from the app). For the app to result in statistically significant increases in PA and reductions to obesity prevalence over 5 years, there needs to be at least 75% of women aware of the app, 75% of those aware of the app download it, and 75% of those who download it opt into the app's push notifications. Even so, the app cannot fully overcome lack of access to recreation centers, and therefore, public health administrators as well as parks and recreation agencies might consider incorporating this type of technology into multilevel interventions that also target the built environment and other social determinants of health.

Acknowledgments

This project was supported by the National Institutes of Health (NIH) Intramural Research Program via grants ZIA HL006168, ZIA HL006225, ZIA HL006252, ZIA MD000010, and ZIA MD000020. The Social Determinants of Obesity and Cardiovascular Risk Laboratory is funded by the Division of Intramural Research at the National Heart, Lung, and Blood Institute (NHLBI) and the Intramural Research Program of the National Institute on Minority Health and Health Disparities (NIMHD). The Socio-Spatial Determinants of Health (SSDH) Laboratory is supported by the Intramural Research Program, NIMHD, NIH, and by the NIH Distinguished Scholars Programs.

This project was also supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) via grant U01HD086861 and 5R01HD086013-02, as well as the NICHD and Office of Behavioral and Social Sciences Research (OBSSR) under award number U54HD070725, the Agency for Healthcare Research and Quality (AHRQ) via grant 1R01HS028165-01, Models of Infectious Disease Agent Study (MIDAS) network under grant R01GM127512, as well as the NIH Common Fund, and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number U54TR004279. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, US Department of Health and Human Services, NHLBI, NIMDH, NICHD, AHRQ, OBSSR, and MIDAS. None of the study sponsors had any role in the study design, collection, analysis, and interpretation of data, writing the report, nor the decision to submit the report for publication. All authors confirm that they had full access to all the data in the study and accept responsibility to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.
[DOCX File , 21 KB-Multimedia Appendix 1]

References

1. ClassPass. URL: https://classpass.com/ [accessed 2022-07-13]
2. The premier fitness membership for studio and online classes. FitReserve. URL: https://www.fitreserve.com/ [accessed 2022-07-13]
3. James DCS, Harville C. Smartphone Usage, Social Media Engagement, and Willingness to Participate in mHealth Weight Management Research Among African American Women. Health Educ Behav 2018 Jun 13;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
4. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
5. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 24;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
6. Estabrooks PA, Lee RE, Gyurcsik NC. Resources for physical activity participation: Does availability and accessibility differ by neighborhood socioeconomic status? ann. behav. med 2003 Apr;25(2):100-104. [doi: 10.1207/s15324796abm2502_05]
7. Reaching Black women across media platforms. Nielsen. URL: https://www.nielsen.com/us/en/insights/article/2017/reaching-black-women-across-media-platforms/ [accessed 2022-07-13]
8. Powell-Wiley TM, Wong MS, Adu-Brimpong J, Brown ST, Hertenstein DL, Zenkov E, et al. Simulating the Impact of Crime on African American Women's Physical Activity and Obesity. Obesity (Silver Spring) 2017 Dec;25(12):2149-2155 [FREE Full text] [doi: 10.1002/oby.22040] [Medline: 29086471]
9. Lee BY, Ferguson MC, Hertenstein DL, Adam A, Zenkov E, Wang PI, et al. Simulating the Impact of Sugar-Sweetened Beverage Warning Labels in Three Cities. Am J Prev Med 2018 Feb;54(2):197-204 [FREE Full text] [doi: 10.1016/j.amepre.2017.11.003] [Medline: 29249555]
10. Hall KD, Butte NF, Swinburn BA, Chow CC. Dynamics of childhood growth and obesity: development and validation of a quantitative mathematical model. The Lancet Diabetes & Endocrinology 2013 Oct;1(2):97-105. [doi: 10.1016/s2213-8587(13)70051-2]

11. Rahmandad H. Human growth and body weight dynamics: an integrative systems model. PLoS One 2014 Dec 5;9(12):e114609 [FREE Full text] [doi: 10.1371/journal.pone.0114609] [Medline: 25479101]
12. WebTrac Facility Search. DC Department of Parks and Recreation. URL: http://app.dpr.dc.gov/dprmap/details.asp?cid=2 [accessed 2019-06-14]
13. Wheaton WD, Cajka JC, Chasteen BM, Wagener DK, Cooley PC, Ganapathi L, et al. Synthesized Population Databases: A US Geospatial Database for Agent-Based Models. Methods Rep RTI Press 2009 May 01;2009(10):905 [FREE Full text] [doi: 10.3768/rtipress.2009.mr.0010.0905] [Medline: 20505787]
14. Adults Health and Wellness. DC Department of Parks and Recreation. URL: https://dpr.dc.gov/page/adults-health-wellness
15. Ries AV, Yan AF, Voorhees CC. The neighborhood recreational environment and physical activity among urban youth: an examination of public and private recreational facilities. J Community Health 2011 Aug 12;36(4):640-649. [doi: 10.1007/s10900-010-9355-1] [Medline: 21225329]
16. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc 1993 Jan;25(1):71-80. [doi: 10.1249/00005768-199301000-00011] [Medline: 8292105]
17. American time use survey: arts activities, [United States], 2003-2017 (ICPSR 36268). National Archive of Data on Arts & Culture. 2017. URL: https://www.icpsr.umich.edu/web/NADAC/studies/36268/versions/V4 [accessed 2022-07-13]
18. Behavioral Risk Factor Surveillance System. DC Health. URL: https://dchealth.dc.gov/service/behavioral-risk-factor-surveillance-system [accessed 2018-01-01]
19. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
20. More than half of U.S. smartphone users ever get news alerts; about half of those who get them click for more information. Pew Research Center. 2016 Sep 07. URL: https://www.pewresearch.org/fact-tank/2016/09/08/more-than-half-of-smartphone-users-get-news-alerts-but-few-get-them-often/ft_16-09-08_newsalerts_main [accessed 2022-07-13]
21. Alpert D. Latest data shows plenty of car-free living in DC. Greater Greater Washington. 2012 Feb 21. URL: https://ggwash.org/view/12290/latest-data-shows-plenty-of-car-free-living-in-dc
22. Smith A. Record shares of Americans now own smartphones, have home broadband. PEW Research Center. 2017 Jan 12. URL: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/
23. Perro J. Mobile Apps: What's A Good Retention Rate? Upland. 2016. URL: https://uplandsoftware.com/localytics/resources/blog/mobile-apps-whats-a-good-retention-rate/
24. Burke M, Brown AL. Distances people walk for transport. Road & Transport Research: A Journal of Australian and New Zealand Research and Practice 2007 Sep;16(3):16-29.
25. Yang Y, Diez-Roux AV. Walking distance by trip purpose and population subgroups. Am J Prev Med 2012 Jul;43(1):11-19 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.015] [Medline: 22704740]
26. Kim SC, Yoon D, Han EK. Antecedents of mobile app usage among smartphone users. Journal of Marketing Communications 2014 Aug 22;22(6):653-670. [doi: 10.1080/13527266.2014.951065]
27. McLean G. Examining the determinants and outcomes of mobile app engagement - A longitudinal perspective. Computers in Human Behavior 2018 Jul;84:392-403. [doi: 10.1016/j.chb.2018.03.015]
28. Michie S, West R. Behaviour change theory and evidence: a presentation to Government. Health Psychology Review 2013 Mar;7(1):1-22. [doi: 10.1080/17437199.2011.649445]
29. Suls J, Mogavero JN, Falzon L, Pescatello LS, Hennessy EA, Davidson KW. Health behaviour change in cardiovascular disease prevention and management: meta-review of behaviour change techniques to affect self-regulation. Health Psychol Rev 2020 Mar 29;14(1):43-65 [FREE Full text] [doi: 10.1080/17437199.2019.1691622] [Medline: 31707938]
30. Cho H, Porras T, Baik D, Beauchemin M, Schnall R. Understanding the predisposing, enabling, and reinforcing factors influencing the use of a mobile-based HIV management app: A real-world usability evaluation. Int J Med Inform 2018 Sep;117:88-95 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.007] [Medline: 30032969]
31. Tong HL, Coiera E, Laranjo L. Using a Mobile Social Networking App to Promote Physical Activity: A Qualitative Study of Users' Perspectives. J Med Internet Res 2018 Dec 21;20(12):e11439 [FREE Full text] [doi: 10.2196/11439] [Medline: 30578201]
32. Moore LV, Diez Roux AV, Evenson KR, McGinn AP, Brines SJ. Availability of recreational resources in minority and low socioeconomic status areas. Am J Prev Med 2008 Jan;34(1):16-22 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.021] [Medline: 18083446]
33. Lee BY, Bartsch SM, Mui Y, Haidari LA, Spiker ML, Gittelsohn J. A systems approach to obesity. Nutr Rev 2017 Jan 03;75(suppl 1):94-106 [FREE Full text] [doi: 10.1093/nutrit/nuw049] [Medline: 28049754]
34. Lee BY, Ferguson MC, Cox SN, Phan PH. Big Data and Systems Methods: The Next Frontier to Tackling the Global Obesity Epidemic. Obesity (Silver Spring) 2021 Feb 09;29(2):263-264 [FREE Full text] [doi: 10.1002/oby.23062] [Medline: 33421308]

Abbreviations

PA: physical activity

Edited by A Mavragani; submitted 21.05.21; peer-reviewed by L Rovniak, A Chwałczyńska; comments to author 30.07.21; revised version received 17.12.21; accepted 30.06.22; published 22.08.22

Please cite as:
Powell-Wiley TM, Martinez MF, Tamura K, Neally SJ, O'Shea KJ, Curlin K, Albarracin Y, Vijayakumar NP, Morgan M, Ortiz-Chaparro E, Bartsch SM, Osei Baah F, Wedlock PT, Ortiz-Whittingham LR, Scannell S, Potharaju KA, Randall S, Solano Gonzales M, Domino M, Ranganath K, Hertenstein D, Syed R, Weatherwax C, Lee BY
The Impact of a Place-Tailored Digital Health App Promoting Exercise Classes on African American Women's Physical Activity and Obesity: Simulation Study
J Med Internet Res 2022;24(8):e30581
URL: https://www.jmir.org/2022/8/e30581
doi: 10.2196/30581
PMID: